

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Docket No.: 200936US0PCT

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF

Patrick M. ALLIEL, et al.

: ART UNIT: 1648

SERIAL NO: 09/719,554

:

FILED: JANUARY 18, 2001

: EXAMINER: CHEN, S.S.

FOR: NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY WITH
HUMAN ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES

CERTIFICATE OF TRANSLATION

COMMISSIONER OF PATENTS
ALEXANDRIA, VIRGINIA 22313

Sir:

Mrs Hélène LEBLOIS-PREHAUD, a translator, residing at

CABINET ORES - 36, rue de St Pétersbourg - 75008 PARIS - FRANCE

hereby states:

- (1) that I am fluent in both the French and English languages;
- (2) that I translated the attached document identified as corresponding to FR 98 07920 filed in France on June 23, 1998, from French to English;
- (3) that the attached English translation is a true and correct translation of FR 98 07920, to the best of my knowledge and belief; and
- (4) that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true and further, that these statements are made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 USC 1001, and that such false statements may jeopardize the validity of the application or any patent issuing thereon.

Date: July 7th, 2004 By: Mrs Hélène LEBLOIS-PREHAUD



INPI

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

P A T E N T

UTILITY CERTIFICATE - CERTIFICATE OF ADDITION

OFFICIAL COPY

The Director-General of the Institut National de la Propriété Industrielle certifies that the attached document is a true copy of an application for industrial property titleright filed at the Institute.

Drawn up in Paris, July 5th, 2004

On behalf of the Director-General of the
Institut National de la Propriété Industrielle
The Patent Department Head

(signature)

Martine PLANCHE

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

REGISTERED OFFICE
26bis, rue de Saint Petersburg
75800 PARIS Cédex 08
Telephone: 33 (1) 53 04 53 04
Fax: 33 (1) 42 93 59 30
www.inpi.fr

INSTITUT NATIONAL DE LA
PROPRIETE INDUSTRIELLE

**26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Telephone: 01 53 04 53 04 Telefax: 01 42 93 59 30**

Intellectual Property Code - Book VI

Cerfa
No. 55-1328

REQUEST FOR GRANT

Confirmation of filing by fax ☐

This form is to be completed in black ink and in block capitals

DB 540a W /170299

This form is to be completed in black ink and in block capitals		DB 540a W /170299						
Reserved for the INPI DATE OF SUBMISSION OF THE DOCUMENTS June 23rd, 1998 NATIONAL REGISTRATION 98 07920 DEPARTMENT OF FILING DATE OF FILING June 23rd, 1998		1. NAME AND ADDRESS OF THE APPLICANT OR THE REPRESENTATIVE TO WHOM THE CORRESPONDENCE IS TO BE ADDRESSED CABINET ORES 6 Avenue de Messine 75008 PARIS No. of permanent power of attorney _____ Correspondent's references _____ Telephone _____ BLOcp598/21FR						
2. APPLICATION <input checked="" type="checkbox"/> patent Nature of the industrial property right <input type="checkbox"/> divisional application <input type="checkbox"/> utility certificate <input type="checkbox"/> conversion of a European patent application <input type="checkbox"/> patent <div style="text-align: center;">→ initial application ↓</div>		<input type="checkbox"/> utility certificate No. _____ date _____ <input type="checkbox"/> deferred <input checked="" type="checkbox"/> immediate <input type="checkbox"/> yes <input type="checkbox"/> no						
Compilation of the search report The applicant, as a physical person, asks to pay the fee by instalments								
Title of the invention (maximum 200 characters) NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY WITH HUMAN ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES								
3. APPLICANT(S) SIREN No. _____ APE-NAF code _____ Name and forenames (underline the surname) or company name INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE - INSERM Nationality/Nationalities French		Legal form Public Establishment						
Full address(es) 101, rue de Tolbiac, 75654 PARIS Cedex 13		Country FRANCE						
If insufficient space, continue on plain paper <input type="checkbox"/>								
4. INVENTOR(S) The inventors are the applicants <input type="checkbox"/> yes <input checked="" type="checkbox"/> no If the answer is no, provide a separate designation								
5. REDUCTION OF THE RATE OF FEES <input type="checkbox"/> requested for the first time <input type="checkbox"/> requested prior to filing; attach copy of the favourable decision								
6. PRIORITY DECLARATION OR APPLICATION FOR THE BENEFIT OF THE FILING DATE OF A PRIOR APPLICATION <table style="width: 100%;"> <tr> <td style="width: 33%;">Country of origin</td> <td style="width: 33%;">Number</td> <td style="width: 33%;">Filing date</td> </tr> <tr> <td></td> <td></td> <td>Nature of the application</td> </tr> </table>			Country of origin	Number	Filing date			Nature of the application
Country of origin	Number	Filing date						
		Nature of the application						
7. DIVISIONS previous to the present application <table style="width: 100%;"> <tr> <td>No.</td> <td>date</td> <td>No.</td> <td>date</td> </tr> </table>			No.	date	No.	date		
No.	date	No.	date					
8. SIGNATURE OF THE APPLICANT OR REPRESENTATIVE (name and capacity of the signatory - registration No.) (signature) Béatrice ORES (No. 92-4046)		SIGNATURE OF THE RECEIVING OFFICIAL SIGNATURE AFTER REGISTRATION OF THE APPLICATION AT THE INPI (illegible signature)						

INPIINSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE**PATENT****UTILITY CERTIFICATE**

Intellectual Property Code - Book VI

Cerfa

N° 11235*02

PATENTS DEPARTMENT

26 bis, rue de Saint-Petersbourg

75800 Paris Cédex 08

Tel: 01 53 04 53 04 Fax: 01 42 93 59 30

DESIGNATION OF THE INVENTOR(S) Page No. 1 / 1

(if the applicant is not the inventor or the sole inventor)

DB 113 W / 260899

NATIONAL REGISTRATION NO.

98 07920

TITLE OF THE INVENTION (200 characters or spaces maximum)NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY WITH HUMAN
ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES**APPLICANT(S):**CABINET ORES
6, avenue de Messine
75008 PARIS**DESIGNATE(S) AS INVENTOR(S):** (indicate name, surnames, adress and underline the name)ALLIEL Patrick M.4, rue Lazare Carnot
92140 CLAMART (FRANCE)PERIN Jean-Pierre182, rue d'Aulnay
92350 LE PLESSIS-ROBINSON (FRANCE)RIEGER François38bis Boulevard de la République
92100 BOULOGNE (FRANCE)NOTA : The name of the inventor can be followed by the name of his membership company when
said membership company is different from that of the applicant or proprietor.**DATE AND SIGNATURE(S)****OF THE APPLICANT(S)****OR OF THE REPRESENTATIVE**

(Name and capacity of the signatory)

Paris, December 21st, 1998

Béatrice ORES (signature)

Representative

N°92-4046

DOCUMENT CONTAINING CORRECTIONS

(FRENCH) PAGE(S) OF THE DESCRIPTION OR OF THE CLAIMS OR SHEET(S) OF DRAWINGS			* R.M.	DATE OF THE CORRESPONDENCE	DATE STAMP OF THE CORRECTOR
Amended	Omitted	Added			
6				12.11.98	16 NOV.1998-SR
73 to 76			X	15.01.99	19 JAN. 1999-SR
73 to 76			X	22 nov 99	EML-06 FEB 2002

* A change made in the wording of the original claims, unless the change derives from the provisions of Article 28 of the decree of 19th September 1979, is indicated by the reference "R.M." (amended claims).

NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY
WITH HUMAN ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES

5 The present invention relates to a novel
nucleic sequence and deduced protein sequence family
with complete or partial human endogenous retroviral
motifs.

The invention also relates to the detection
and/or use of said nucleic sequences and of said
10 corresponding protein sequences in the context of
diagnostic, prophylactic and therapeutic applications,
in particular for neuropathological conditions with an
autoimmune component such as multiple sclerosis.

The invention also relates to the production of
15 antisense double-stranded and single-stranded nucleic
probes, of ribozymes, capable of modulating viral
replication (T.R. Cech, *Science*, 1987, **236**, 1532-1539;
R.H. Symons, *Trends Biochem. Sci.*, 1989, **14**, 445-450)
of the corresponding recombinant molecules, and
20 associated antibodies.

Retroviruses are viruses which replicate solely
by using the opposite route to the conventional
processing of genetic information. This process, called
reverse transcription, is mediated by an RNA dependent
25 DNA polymerase or reverse transcriptase, encoded by the
pol gene. Retroviruses also encode at least two
additional genes. The *gag* gene encodes the proteins of
the skeleton, matrix, nucleocapsid and capsid. The *env*
gene encodes the envelope glycoproteins. Retroviral
30 transcription is regulated by promoter regions or
"enhancers" situated in highly repeated regions or LTR
(*Long Terminal Repeat*) and which are present at both
ends of the retroviral genome.

During the infection of a cell, polymerase
35 makes a DNA copy of the RNA genome; this copy may then
integrate into the human genome. Retroviruses do not
kill the cells which they infect, but on the contrary
often enhance their rate of growth. Retroviruses can
infect germ cells or embryos at an early stage; they

can, under these conditions, integrate the germ line and be transmitted by vertical Mendelian transmission, which constitutes the closest relationship between a host and its parasite. These endogenous viruses can degenerate during generations of the host organism and lose their initial properties. However, some of them may conserve all or part of their properties or of the properties of their constituent motifs, or acquire novel functional properties having an advantage for the host organism, which would explain the preservation of their sequence.

The existence of endogenous motifs having long open reading frames and/or subjected to a strong selection pressure can therefore be an indication of a preserved or acquired biological function, which may correspond to a benefit for the host organism. These retroviral sequences can also undergo, over the generations, discrete modifications which will be able to trigger some of their potentials and generate or promote pathological processes. It has recently appeared necessary to carry out a review and to identify these sequences so as to be able to evaluate their functional impact.

Human endogenous retroviral sequences or HERVs represent a substantial part of the human genome. These retroviral regions exist in several forms:

- complete endogenous retroviral structures combining *gag*, *pol* and *env* motifs, flanked by repeat nucleic sequences which exhibit a significant analogy with the LTR-*gag-pol-env*-LTR structure of infectious retroviruses,

- truncated retroviral sequences; for example the retrotransposons lack their *env* domain and the retroposons do not possess the *env* and LTR regions.

Up until now, the study of these regions of the genome has been neglected in humans for essentially two reasons:

- the existence of insertions/deletions which can shift the reading frame and of mutations which modify the sequence. These modifications cause impairment of the structure and consequently of the biological function of these motifs,

- the absence of confirmed associations with human pathological conditions.

The recent knowledge of fragments which are significantly representative of the human genome and an orientation of research studies toward a study of structure/function of endogenous retroviral motifs have made it possible to specify the importance of these regions. The involvement of truncated or complete endogenous sequences in pathological conditions in animals is documented; for example their association with tumor processes has been clearly demonstrated (S.K. Chattopadhyay et al., 1982, *Nature*, **295**, 25-31). Research aimed at specifying the association or the influence of HERVs in human pathological conditions is now therefore justified.

A classification of the HERV elements has been proposed (Tönjes R.R. et al., *AIDS & Hum. Retroviral.*, 1996, **13**, p261-p267; A.M. Krieg et al., *FASEB J.*, 1992, **6**, 2537-2544). It is based on a homology of these sequences with retroviruses isolated in animals, with the aid of heterologous retroviral probes. Indeed, in general, the HERVs exhibit relatively little homology with known human infectious retroviruses.

The class I families exhibit a sequence homology with the type C mammalian retroviruses; there may be mentioned in particular the ERI superfamily, close to the MuLV virus (*murine leukemia virus*) and to the BaEV virus (*baboon endogenous virus*).

The class II families exhibit a sequence homology with the type B mammalian retroviruses such as MMTV (*mouse mammary tumor virus*) or the type D retroviruses such as SRV (*squirrel monkey retrovirus*).

Other families have also been described; among these, there may be mentioned HERVs which exceptionally exhibit partial homology with HTLV-1 (RTVL-H) or primate viruses; HRES-1, for example, exhibits sequence
5 homology with HTLVs.

Programmes for very large sequencing of the human genome now make it possible to have available a significant number of novel retroviral sequences. The use of data processing software packages makes it
10 possible to identify and analyse these genes. In this context, a systematic search relating to the entire information available to date has been initiated in order to identify novel human endogenous retroviral sequences as a function of certain analytical criteria:

15 - presence of long open reading frames conserved during evolution of the host organism and which may suggest a biological function,

- analogy with sequences already characterized outside or inside the retrovirus domain,

20 - location in regions of susceptibility for certain pathological conditions or close to essential genes, for example in the cancer domain, regulation of the immune system or in certain neuropathological conditions.

25 The work carried out by the inventors on sequence databases allowed them to identify a set of endogenous retroviral sequences or motifs whose normal or pathological expression can promote or disrupt a protective effect in relation to pathological
30 processes, or play a role in the onset or worsening of pathological conditions.

The subject of the present invention is a purified nucleic acid fragment, characterized in that it comprises all or part of a sequence encoding a human
35 endogenous retroviral sequence, which has at least env-type retroviral motifs, corresponding to the sequence SEQ ID NO: 1 or to a sequence exhibiting a level of homology with said sequence SEQ ID NO: 1 greater than

or equal to 80% on more than 190 nucleotides or greater than or equal to 70% on more than 600 nucleotides for the *env*-type domains.

5 The expression homologous sequence is understood to mean both a sequence which exhibits complete or partial identity with the abovementioned sequence SEQ ID NO: 1 and a sequence which exhibits partial similarity with said sequence SEQ ID NO: 1.

10 According to an advantageous embodiment of said fragment, it has retroviral motifs corresponding to an *env* domain and corresponding to the sequence SEQ ID NO: 1 and retroviral motifs corresponding to a *gag* domain and corresponding to the sequence SEQ ID NO: 2 or to a sequence exhibiting a level of
15 homology greater than or equal to 80% on more than 190 nucleotides or greater than or equal to 70% on more than 600 nucleotides for the *env*-type domains and a level of homology greater than or equal to 90% on more than 700 nucleotides or greater than or equal to 70% on
20 more than 1 200 nucleotides for the *gag*-type domains, said motifs having no insertion or deletion of more than 200 nucleotides.

Said fragments constitute a novel family of human endogenous retroviral sequences (HERV-7q family)
25 which exhibits sequence homology with the MSRV retroviruses, as described in International Application WO 97/06260; said fragments according to the present invention have:

- two repeat nucleotide motifs of 711 bp
30 (Figure 3), having characteristic signals identified in LTRs (*Long Terminal Repeats*): transcription promoters of the TATAA or CCAAT box type. These repeat domains delimit three deduced motifs of the *gag*, *pol* and *env* type (Figure 2),

35 - an *env*-type motif (positions 6965 nt - 9550 nt on the sequence SEQ ID NO: 3) which contains a long open reading frame of 1 620 nucleotides (positions 7874-9493 of the sequence ID NO: 3) encoding a protein

having an unpublished sequence of 540 amino acids (Figure 4) and underlined fragment in SEQ ID NO: 27. There is present inside the transmembrane domain of this env domain a peptide motif of the CKS-25/CKS-17 type (Figure 5), recognized as having immunosuppressive functions on the host lymphocytic cells (M. Mitani et al., 1987, *Proc. Natl. Acad. Sci. USA*, **84**, 237-240). A zinc finger type domain **HX₃₋₄HX₂₂₋₃₃CX₂C** (Kulkolski et al., 1992, *Mol. Cell. Biol.*, **12**, 2331-2338), which is present in integrase-type domains is identified in another reading frame. This particular env domain signatures the characteristic of novel endogenous retroviral motifs,

the motif (positions 3065 nt - 4390 nt on the sequence SEQ ID NO: 3) of the gag type encoding protein motifs according to Figure 6 (SEQ ID NO: 51) (positions 3118-4198 of SEQ ID NO: 3) was identified by virtue of analogies with known gag domains. The region of major homology **QX₃EX₇R** is for example present (Benit et al., 1997, *J. Virol.*, **71**, 5652-5657). The nucleic acid binding motif **CX₂CX₃₋₄HX₄C**, situated at the C-terminal position, is identified in another reading frame (Covey et al., 1986, *Nucleic Acids Res.*, **14**, 623-633). Upstream of the gag domain, a motif of 182 nucleotides is detected which is repeated twice (Figure 1),

- the pol domain exhibits the conventional consensus of a retrovirus pol region at the level of the protease, reverse transcriptase and RNase H domains. A motif close to the consensus **LLDTGA** is found in pol (Weber et al., 1988, *Science*, **243**, 928-931). The motifs **D** and **AF**, **LPQ** and **SP**, and **YVDD** (Xiong and Eickbush, 1990, *EMBO J.*, **9**, 3353-3362) are respectively found in the 3rd, 4th and 5th homology boxes. The motifs **YTDGSS** and **TDS** are present in the RNase H region,

- the gag and pol regions could be considered as being joined with a passage from the gag region to the pol region by a reading frame shift.

The present invention includes the sequences belonging to the HERV-7q family as defined above (presence of the SEQ ID NO: 1 sequence or of a homologous sequence or presence of both the sequences
5 SEQ ID NO: 1 and SEQ ID NO: 2) and in particular the sequences SEQ ID NO: 3-21; it also includes the complementary nucleic sequences and the reverse sequences complementary to the preceding sequences as well as fragments derived from the coding regions of
10 the preceding sequences corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences (SEQ ID NO: 30-50).

These various fragments may be advantageously used as primers or as probes ; they hybridize
15 specifically to a sequence of the HERV-7q family.

Among these fragments, the following fragments may be preferably mentioned:

- a fragment of 182 nucleotides, repeated twice, situated upstream of the *gag* domain at positions
20 2502-2611/2613-2865 of SEQ ID NO: 3:

Primers and probes specific for the *gag* region

- a sense primer G1F located in the region upstream of the *gag* domain of *HERV-7q*:
5'GGACCATAGAGGACACTCCAGGACTA3' (SEQ ID NO: 30);

25 - an antisense primer G1R located in the terminal 3' region of the *gag* domain:
5'CCTCAGTCCTGCTGCTGGATCATCT3' (SEQ ID NO: 31)

- the fragment of 1505 nt amplified by the pair G1F-G1R is used in order to generate the probes capable
30 of hybridizing the various PCR amplification products:

- a nested sense primer G2F: (SEQ ID NO: 32)
5'CCTCCAAGCAGTGGGAGGAAGAGAATT3'

- a nested antisense primer G2R: (SEQ ID NO: 33)
5'CCTTCCCTGTGTTATTGTGGACATCATT3'

35 - a nested sense primer G4F: (SEQ ID NO: 34)
5'GGAAGAAGTCTATGAATTATTCAATGATGT3'

- a nested sense primer G3F: (SEQ ID NO: 35)
5'GGGACACAGAATCAGAACATGGAGATT3'

- a nested antisense primer G4R: (SEQ ID NO: 36)

5'GCCTTCAGAAGAGTCAGGTGACAGAGA3'

- a nested antisense primer G5R: (SEQ ID NO: 37)

5'GAGCCTCCAAAGTCCACTTGCCTGA3'

5 Primers and probes specific for the env region

- a sens primer E1F: (SEQ ID NO: 38)

5'GATTTTCAGTATCTACTAGTCTGGGTAGAT3'

- an antisense primer E1R: (SEQ ID NO: 39)

5'CTAGGAAATCCAGCTAGTCCTGTCTCA3'

10 - the fragment of 2529 nt, amplified by the pair
of primers E1F-E1R, is used to generate the probes
capable of hybridizing the various PCR amplification
products:

- a sense primer E2F: (SEQ ID NO: 40)

15 5'CCAAGACAGCCAACTTAGTTGCAGACAT3'

- an antisense primer E2R: (SEQ ID NO: 41)

5'GGACGCTGCATTCTCCATAGAACTCTT3'

- a sense primer E3F: (SEQ ID NO: 42)

5'GCAATACTACATACACAACCAACTCCCAA3'

20 - an antisense primer E3R: (SEQ ID NO: 43)

5'GGGGGAGGCATATCCAACAGTTAGTA3'

- a sense primer E4F: (SEQ ID NO: 44)

5'CCATCTACACTGAACAAGATTTATACACTT3'

- an antisense primer E4R: (SEQ ID NO: 45)

25 5'AATGCCAGTACCTAGTGCACCTAGCACT3'

- a sense primer E5F: (SEQ ID NO: 46)

5'CGAATACAACGTAGAGCAGAGGAGCTTCGAA3'

- a sense primer E6F: (SEQ ID NO: 47)

5'AGCCCAAGATGCAGTCCAAGACTAAGAT3'

30 - a primer E5R: (SEQ ID NO: 48)

5'GCGTAGTAGAGGTTGTGCAGCTGAGAT3'

- a primer ExF: (SEQ ID NO: 49)

CCCTTACCAAGAGTTTCTATGGAGAAT

- a primer ExR: (SEQ ID NO: 50)

35 ACCGCTCTAACTGCTTCCTGCTGAATT

All the oligonucleotides are designed to be able
to generate a sense primer and an antisense primer by a
shift in the sequence of the reference primer of 1 to 7

nucleotides toward the 5' side or toward the 3' side; the modification of the sequence may cause a modification of the size of the primer of 1 to 7 nucleotides depending on the cases. The primers chosen
5 may be optimized depending on the cases by shortening or extension affecting 1 to 9 nucleotides.

Preferably, the hybridization, cloning, subcloning, production, preparation and analysis of the nucleic acids, peptides and antibodies, the sequencing
10 of the nucleic acids and peptides, the *in situ* hybridization and the immunohistochemistry are carried out under the conditions described in the following books:

- Current Protocols in Molecular Biology, Eds.
15 F.M. Ausubel, R. Brent & R.E. Kingston et al. Green Publishing associates and Wiley Interscience.

- Molecular Cloning: a laboratory manual. Eds. J. Sambrook, E.F. Fritsch & T. Maniatis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

20 - The Practical Approach series. Eds. D. Rickwood & B.D. Ames, IRL Press and Oxford University Press. In particular antibodies I & II; DNA cloning I, II, III; Nucleic acid and protein sequence analysis; Nucleic acid hybridization; Nucleic acid
25 sequencing; Oligonucleotide synthesis; Protein purification applications; Protein purification methods; Protein sequencing; Transcription and translation; Gels electrophoresis of nucleic acids; Gels electrophoresis of proteins; Genome analysis; HPLC
30 of macromolecules; Human genetic diseases; Microcomputing in biology; Molecular neurobiology; Mutagenicity testing; Essential molecular biology I & II.

- Proteome research: New frontiers in
35 functional genomics, Eds. M.R. Wilkins et al., Springer.

The human endogenous retroviral sequence (SEQ ID NO: 3) situated on the long arm of chromosome 7

corresponds to the HERV-7q sequence; it has 10.5 kb (Figs. 1 and 2) and satisfies the criteria defined above.

5 The search for domains exhibiting total or partial similarity with the *gag* and *env* regions of HERV-7q resulted in the identification of novel endogenous retroviral sequences. These sequences may have the structure of a complete endogenous retrovirus such as the endogenous retroviral sequence situated
10 close to the gene for the alpha and delta subunits of the T cell receptor, and consequently called HERV-TcR; by way of example, Figure 7 shows the comparison of the nucleic alignments of the respective *gag* domains of HERV-7q and HERV-TcR (sequence HG12, SEQ ID NO: 18).
15 Partial retroviral structures also exist. These retroviral domains, similar to HERV-7q, are identified in independent nucleic sequences as shown by their chromosomal location. Nucleic motifs (called here HEx or HGx, and analogous to *env* or *gag* type domains,
20 respectively) resembling the *env* or *gag* domains of HERV-7q were found, with the aid of the abovementioned databases:

- HE2: chromosome 17 (SEQ ID NO: 4),
- HE3 and HG3: chromosome 6 (SEQ ID NO: 5 and 6),
- 25 - HE4: chromosome X (SEQ ID NO: 7),
- HE5: chromosome X q22 (SEQ ID NO: 8),
- HE6 and HG6: chromosome 1 q23.3-q24.3 (SEQ ID NO: 9 and 10),
- HE7: chromosome 7 p15 (SEQ ID NO: 11),
- 30 - HE8 and HG8: chromosome 19 (SEQ ID NO: 12 and 13),
- HE9: chromosome X (SEQ ID NO: 14),
- HE10: chromosome X q13.1-21.1 (SEQ ID NO: 15),
- HE11 and HG11: chromosome 7 q21-22 (SEQ ID NO: 16 and 17),
- 35 - HE12 and HG12, in HERV-TcR: chromosome 14 q11.2 (SEQ ID NO: 18 and 19),

The alignments of the *env* (Fig. 8) and *gag* (Fig. 9) domains explain the levels of homology observed between the sequences described above and the homologous sequences in HERV-7q. The analogies can
5 extend to the flanking retroviral motifs.

Analysis of the sequence tags available in databases shows that transcripts belonging to some members of this family, in particular HERV-7q, are essentially expressed in tissues of foetal or placental
10 origin.

Polypeptide sequences generated by these transcripts can therefore be potentially produced and biological functions or activities can be envisaged, by analogy with biologically active polypeptides of viral
15 or retroviral origin; for example, the peptide motifs of the CKS-17 type (Fig. 5) or CKS-25 type (Huang S.S. and Huang J.S., J. Biol. Chem. 1998, 273, 4815-4818) which have immuno-modulatory functions on the lymphocytic host cells. The differences in sequence
20 which are observed and possible normal or pathological modifications are in particular responsible for modulation of the function.

HERV-7q represents the paradigm of the novel family of human endogenous retroviral sequences or of
25 endogenous retroviral motifs.

HERV-7q and some of the endogenous retroviral sequences belonging to its family have a *pol*-type domain analogous to *pol*-type retroviral sequences such as for example the *pol* region identified in the MSRV
30 retrovirus associated with multiple sclerosis and described by H. Perron et al. (1997, *Proc. Natl. Acad. Sci. USA*, **94**, 7583-7588; International Application PCT WO 97/06260).

However, the sequences according to the present
35 invention are distinguishable from the infectious exogenous retroviral sequences analogous to MSRV previously described in that the *gag* and *env* sequences according to the invention are significantly different

according to the criteria defined above and as a function of certain specific characteristics, for example the long open reading frame of the *env* domain of HERV-7q; they would be able to allow the signaturing of a pathological condition when they have insertions, deletions, reading frame shifts or mutations.

Indeed, the differences observed between the human sequences of the HERV-7q type, which are isolated from individuals reputed to be normal, and the sequences derived from some samples of pathological origin are not randomly distributed. Comparisons carried out between the *gag* region obtained from infectious retroviral particles (EMBL accession No.: A60168, A60200, A60201, A60171 and the like) and the corresponding *gag* sequence of HERV-7q (Fig. 9), make it possible to observe that the mutations preferably affect non-sense codons. For example, two non-sense codons in HERV-7q are replaced by an arginine codon in A60200, which makes it possible to obtain a deduced sequence of 109 amino acids for HERV-7q and of 166 amino acids for A60200. The base changes consequently make it possible to extend the reading frame and to potentially encode larger sized polypeptide structures (Figure 10).

Likewise, an *env*-type sequence obtained from infectious retroviral particles exhibits a significant analogy with the *env* domain of HERV-7q (Figure 11). These marked analogies between exogenous and endogenous retroviral sequences could be responsible for the triggering or worsening of certain pathological processes, in particular certain autoimmune diseases such as multiple sclerosis. In this regard, it is possible to note that certain endogenous retroviral sequences described in the invention are situated close to or in regions reputed to exhibit susceptibility for multiple sclerosis: for example HERV-7q and the 7q21-22 region of chromosome 7, likewise for HE12 and HG12 in HERV-TcR and the region of the gene encoding the alpha

and delta chains of the T cell receptor, HE2 and chromosome 17, or HE3, and HG3 and chromosome 6

No significant homology is observed with endogenous retroviral sequences already described; on the other hand, a limited homology may be noted, and in any case said homology is less than the criteria defined according to the invention between the env domains of the sequence HERV-7q (SEQ ID NO: 1) and the sequence HERV-9 (Figure 12). Figure 13 shows extensive homologies between the sequence HERV-7q with an exogenous retroviral sequence (accession No. EMBL: A60170).

The human endogenous retroviral sequences belonging to the HERV-7q family can protect against attacks linked to the environment or can be beneficial for the individual. This beneficial effect could be one of the possible reasons for the selection pressure exerted on some of these sequences and the potentially functional character of the deduced protein structures identified: for example the long open reading frame capable of encoding a novel protein and corresponding to the env domain of HERV-7q.

The human endogenous retroviral sequences belonging to the HERV-7q family could be associated, for example, with pathological conditions related to processes linked to cancer, to neuropathological conditions with an autoimmune component or to any other pathological process in association or otherwise with endogenous or exogenous viruses or retroviruses. Their action could be related to the outbreak, the worsening, the modification of the time of appearance or the protection against the disease.

In the context of application to autoimmune pathological conditions (such as for example lupus, Sjögren's syndrome, rheumatoid arthritis, multiple sclerosis and the like), significant analogies may be detected between the endogenous retroviral motifs identified and motifs found in retroviral structures

characterized in patients with autoimmune pathological conditions such as multiple sclerosis; for example, fragments of *gag* domain (recently available in databases) obtained from infectious retroviral particles or the complete sequence of the *pol* domain corresponding to the MSRV virus associated with multiple sclerosis. These retroviral motifs possess significant analogies with homologous endogenous sequences of the HERV-7q type, which makes it possible to envisage direct or indirect association with pathological processes, including multiple sclerosis, in association or otherwise with MSRV.

The presence of some sequences or motifs can be observed in multiple sclerosis susceptibility regions : for example, the sequences HE 11 and HG 11 around the region 7q 21-22 or moreover HE 4, HE 5, HE 6, HE 9, HE 10 or HG 10 on the X chromosom are located in or nearness chromosomic regions regularly associated with multiples sclerosis susceptibility genes. These sequences should provide means for the localisation or identification of predisposition genes.

The importance of these sequences goes beyond the context of autoimmune diseases. Apart from the general importance of retroviral motifs in the triggering or worsening of a tumor process, which is well established in particular in murine models (H. Fan in *The retroviridae*, 1994, ed. J.A. Levy, Plenum, New York, p. 313-353), these sequences could be present close to or inside important genes and could alter the expression thereof: for example HERV-TcR and the genes for the alpha and delta subunits of the receptor for the T cells involved in disruptions of the immune system. The subject of the invention is also transcripts generated from the abovementioned sequences as well as those optionally exhibiting modifications in the reference sequences described in the invention when they are expressed in certain patients.

Indeed, the systems for regulating the the expression of the retroviral proteins of HERV-7q, which are present in the LTR type motifs, could influence the expression of genes situated in the close or distant
5 chromosomal vicinity and could induce disruptions of an immunological and/or neurological character. For example, the endogenous retroviral sequence HERV-TcR exists in the immediate vicinity of the genes for the alpha and delta subunits of the T cell receptor
10 previously described. The LTR-type motifs could also encode superantigens (Acha-Orbea and Palmer, 1991, *Immunol. Today*, **12**, 356-361). In general, retroviral proteins of the HERV-7q or related type, or their truncated or partial forms could be involved in
15 cytotoxicity or superantigenicity phenomena, such as for example those derived from the long open reading frame identified in the env domain (Figure 4).

In this regard, it is possible to note that retroviral motifs derived from defective regions are
20 capable of having biological functions; for example, the envelope protein p15E, derived from defective retroviral motifs, possesses an anti-inflammatory and immunosuppressive activity (Snyderman and Ciancolo, 1984, *Immunol. Today*, **5**, 240-244).

25 These structures are probably capable of causing breaks or of amplifying deregulations in the immune defense processes. Some of the motifs of the gag, env and LTR-type domains may be associated with a particular function or may contribute to the normal or
30 pathological function of the flanking domains. Recombinations with an element of exogenous, retroviral origin or otherwise can give rise to the production of nucleic or protein motifs which could either protect or trigger or promote or worsen a pathological condition.
35 Likewise, a retroviral structure containing endogenous retroviral elements according to the invention would be capable of causing a pathological process after passing through an exogenous transient cycle followed by

reintegration into a sensitive or critical region of the human genome.

Likewise, the combination of motifs belonging to the HERV-7q family, or of elements induced by motifs belonging to the HERV-7q family, with motifs of exogenous origin or induced exogenously would be capable of triggering or worsening a pathological process or on the contrary of promoting protection or partial remission or a complete and permanent cure.

The detection made possible of the HERV-7q type domains suggests possible applications at the prophylactic, prognostic and diagnostic level; for example, immunological approaches or gene amplification, which make it possible to compare normal individuals serving as reference with patients, would be capable of promoting screening, of improving early detection of the outbreak of the disease and/or of monitoring the progression of a pathological condition in patients which may exhibit a susceptibility or in whom there has been an outbreak of the disease or in individuals considered to be normal, based on current clinical criteria.

The specific nucleic and immunological probes, as defined, in the present invention are capable of promoting the identification and detection of motifs which are abnormally expressed in the context of pathological conditions associated with cancer, or of neuropathological conditions, in particular autoimmune pathological conditions, at the forefront of which is multiple sclerosis.

Therapeutic strategies may be envisaged by using some of the nucleic sequences contained in HERV-7q and the sequences of the same family or deduced polypeptide structures or by the use of peptides or proteins, or of specific antibodies. The subject of the present invention is also hybrid nucleic sequences, characterized in that they comprise sequences or motifs belonging to the HERV-7q family, or of elements induced by motifs belonging to the HERV-7q

family, with motifs of exogenous origin or induced exogenously (exogenous retroviral sequences); such hybrid sequences are probably capable of triggering or worsening a pathological process or on the contrary of promoting protection or partial remission or a complete and permanent cure.

The subject of the present invention is also a diagnostic reagent for the differential detection of complete or partial human endogenous nucleic sequences, having retroviral motifs, selected from the sequences SEQ ID NO: 1 and/or SEQ ID NO: 2, characterized in that it is selected from the group consisting of the sequences SEQ ID NO: 1-50, the complementary nucleic sequences and the reverse sequences complementary to the preceding sequences, of nucleotide fragments capable of defining or of identifying the sequences SEQ ID NO: 1 and/or SEQ ID NO: 2 and any flanking sequence or any sequence overlapping them as well as of fragments derived from the coding regions of the sequences SEQ ID NO: 1-24 corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences, optionally labeled with an appropriate marker.

The sequences of the nucleic, ribonucleic and oligonucleotide probes used will be chosen from the *env* and *gag* regions or their flanking regions; for example the oligonucleotide primers for HERV-7q will be chosen from the regions situated between nucleotides 3065 and 4390, nucleotides 6965 and 9550 as well as from any adjacent sequence (upstream or downstream) capable of allowing specific amplification (Figure 1).

Among the appropriate markers, there may be mentioned radioactive isotopes, enzymes, fluorochromes, chemical markers (biotin), haptens (digoxigenin) and antibodies or appropriate base analogues.

Preferably:

- said reagent is selected from the sequences
SEQ ID NO: 30-50 and is capable of being used as a
primer,

5 - said reagent is selected from the following
sequences:

 a fragment of 1505 nt amplified by the
pair of primers SEQ ID NO: 30 and SEQ ID NO: 31
(primers G1F and G1R),

 a fragment of 2529 nt amplified by the
10 pair of primers SEQ ID NO: 38 and SEQ ID NO: 39
(primers E1F and E1R)

 and is capable of being used as a probe.

The subject of the present invention is also a
method for the rapid and differential detection of the
15 endogenous retroviral nucleic sequences of the env or
env and gag type, their normal or pathological
variants, by hybridization and/or gene amplification,
carried out using a biological sample, which method is
characterized in that it comprises:

20 (a) a step in which a biological sample to
be analysed is brought into contact with at least one
probe as defined above, and

 (b) a step in which the product(s) resulting
from the nucleotide sequence-probe interaction is
25 detected by any appropriate means.

In accordance with said method, it may
comprise:

* prior to step (a):

30 . a step of preparing the relevant biological
tissue or fluid,

 . a step of extracting the nucleic acid to be
detected, and

 . at least one gene amplification cycle, and

* subsequent to step (b):

35 . a step of comparing the nucleic sequences
obtained in said biological sample with the human
endogenous retroviral sequences according to the
invention by any appropriate means and in particular by

sequencing, Southern blotting, restriction cleavage, SSCP or any other method which makes it possible to identify an insertion or a deletion or a single mutation between the various sequences compared.

5 In accordance with the invention, the human endogenous retroviral sequences according to the invention are thus compared with the nucleic sequences present in the biological sample to be analysed and allow the detection of homologous sequences from
10 patients suffering from pathological conditions likely to involve a modification of their genome.

Advantageously, said gene comparisons are carried out using genomic DNA obtained from control individuals and from patients.

15 A conventional gene amplification by PCR will be carried out with the aid of 5'-sense and 3'-antisense primers delimiting or comprising the zone to be studied (*env* zone or *gag* zone).

Also advantageously, the sequences of the
20 nucleic, ribonucleic and oligonucleotide probes used are chosen from the *env* and *gag* regions or their flanking regions; for example the oligonucleotides which are primers for HERV-7q will be chosen from the regions situated between nucleotides 3065 and 4390 and
25 nucleotides 6965 and 9550, and from any adjacent sequence (upstream or downstream) capable of allowing specific amplification (Figure 1), as specified above. They are preferably selected from the group consisting of

30 a fragment of 1505 nt amplified by the pair of primers SEQ ID NO: 30 and SEQ ID NO: 31 (primers G1F and G1R),

 a fragment of 2529 nt amplified by the pair of primers SEQ ID NO: 38 and SEQ ID NO: 39 (primers E1F
35 and E1R).

The gene amplification step is in particular carried out with the aid of one of the following gene

amplification techniques: amplification using
Q β -replicase, PCR, LCR, ERA, CPR or SDA.

5 The subject of the present invention is also a
method of detecting transcripts as defined above,
characterized in that it comprises:

- collecting messenger RNAs obtained from
control biological samples (biological tissues, cells
or fluids) and from a similar sample collected from
10 patients, and

- the qualitative and/or quantitative analysis
of said mRNAs by *in situ* hybridization, by dot-blot,
Northern blotting, RNase mapping or RT-PCR, with the
aid of a diagnostic reagent as defined above.

15 The subject of the present invention is also
products of translation, characterized in that they are
encoded by a nucleotide sequence as defined above.

The subject of the present invention is also a
peptide, characterized in that it is capable of being
20 expressed with the aid of a nucleotide sequence
selected from the group consisting of the sequences
SEQ ID NO: 1-24, as defined above.

Said peptide also includes the derived peptides
or polypeptides comprising between 5 and 540 amino
25 acids (SEQ ID NO: 25-29 and SEQ ID NO: 51 and their
fragments of at least 5 amino acids). Said peptides are
translated from the above defined nucleotide sequences,
according to the combination offered by usage of the
different possible open reading frames.

30 According to an advantageous embodiment of said
peptides they are in particular selected from the
sequences SEQ ID NO: 25-29 and SEQ ID NO: 51.

According to another advantageous embodiment of
said peptides, they are obtained from nucleic sequences
35 as defined above, in which at least one non-sense codon
may be replaced with a codon encoding one of the
following amino acids: Phe (F), Leu (L), Ser (S), Tyr

(Y), Cys (C), Trp (W), Gln (Q), Arg (R), Lys (K), Glu (E) or Gly (G).

The invention thus includes the deduced peptides or the deduced proteins corresponding to all or part of the nucleic sequences described in the invention, and optionally exhibiting modifications with the reference sequences described in the invention, when they are expressed in some patients. In particular, the invention includes the complete or partial sequences obtained according to the 3 sense reading frames and the 3 reverse and complementary reading frames (SEQ ID NO: 22-24).

Advantageously, the env protein of HERV-7q of the invention has :

- N-glycosylation sites. The glycosylation of the envelope proteins of retroviruses appears to be directly associated with their functional properties, for example by influencing the number of determinants available in the T cells or by promoting recognition of antigens by the T cells. Glycosylation could play a role in the outbreak or the spread of a pathological condition with an autoimmune component. The glycosylations are necessary for maintaining the conformation of certain epitopes, in particular during the production of a recombinant envelope protein so as to develop a diagnostic reagent and to promote the efficacy of a possible vaccine. Positions 171, 210, 216, 236, 244, 283 and 411. Expected number at random: 3.2

- prenylation sites. Prenylation is an essential mechanism for attachment to the cell membrane and for the targeting of certain proteins. This targeting process could be essential for the production of specific therapeutic agents capable of interfering with the production and regulation of the traffic of cellular complexes calling into play proteins involved in the cell interactions, growth and movement. Positions 188 and 290. Expected number at random: 1.8

- targeting sites in the endoplasmic reticulum.

These sites could make it possible to bring about the targeting toward the endoplasmic reticulum in order to carry out the modifications necessary for promoting
5 membrane crossing. Positions 353 and 431. Expected number at random: 0.2. Said peptides or proteins can advantageously show biological properties.

The protein products generated by the
10 endogenous retroviral sequences or produced in parallel may be advantageously characterized by micro-methods of analysis and quantification of peptides and proteins: HPLC/FPLC or equivalent, capillary electrophoresis or equivalent, microsequencing techniques (Edman method or
15 equivalent, mass spectrometry and the like).

The subject of the invention is also antibodies directed against one or more of the peptides described above and their use either for carrying out a method, in particular a differential method, of *in vitro*
20 detection of the presence of such a sequence in an individual.

Said antibodies are advantageously polyclonal or monoclonal antibodies obtained by an immunological reaction from a human, mammalian or avian organism or
25 other species toward the proteins, as defined above.

The subject of the present invention is a method for the differential immunological screening of normal or pathological human endogenous retroviral sequences of the HERV-7q family, characterized in that
30 it comprises bringing a biological sample into contact with an antibody according to the invention, the reading of the result being visualized by an appropriate means, in particular EIA, ELISA, RIA, fluorescence.

35 By way of illustration, such an *in vitro* diagnostic method according to the invention comprises bringing a biological sample collected from a patient into contact with antibodies according to the invention and detecting with the aid of any appropriate method,

in particular with the aid of labeled anti-immunoglobulins, the immunological complexes formed between the proteins produced normally or pathologically and the antibodies.

5 Monoclonal or polyclonal antibodies, produced from antigens corresponding to synthetic peptides, or recombinant polypeptide or proteins make it possible to monitor the expression of the peptides or proteins produced normally or pathologically. The analysis is
10 preferably carried out by ELISA or equivalent, Western blotting or equivalent, or by immunohistochemistry.

 The peptides or proteins, derived from the endogenous retroviral sequences or whose expression is associated with the expression of these endogenous
15 retroviral sequences, are tested for and identified.

 The subject of the present invention is also a method for the identification and detection of endogenous retroviral motifs which are abnormally expressed in the context of pathological conditions
20 associated with cancer, or of neuropathological conditions, in particular autoimmune neuropathological conditions, at the forefront of which is multiple sclerosis, characterized in that it comprises the comparative analysis of the sequences extracted from a
25 biological sample and the sequences according to the invention.

 The subject of the present invention is also the application of the nucleic sequences or of the protein sequences according to the invention to the
30 diagnosis of, to the prognosis of, to the evaluation of genetic susceptibility to, any induced, congenital or acquired human diseases, in particular those with cancerous, autoimmune and/or neurological components, such as multiple sclerosis, the associated syndromes
35 and the neurodegenerative diseases in which all or part of the nucleic sequences according to the invention and related endogenous or exogenous forms are involved.

The subject of the present invention is also hybrid nucleic sequences, characterized in that they comprise nucleic sequences or motifs according to the invention, combined with sequences or motifs of endogenous origin or of exogenous origin or induced exogenously.

The subject of the present invention is, in addition, a recombinant cloning or expression vector, characterized in that it comprises a nucleic sequence in accordance with the invention.

In addition to the preceding arrangements, the invention also comprises other arrangements which will emerge from the description which follows, which refers to exemplary embodiments of the method which is the subject of the present invention as well as to the appended drawings, in which:

- Figure 1. Human nucleic sequence HERV-7q, whose analysis and treatment make it possible to characterize a novel endogenous retroviral structure. The repeat nucleic regions of type R1 and R2 and the *gag*, *pol* and *env* domains are underlined. The *gag* and *env* type domains are in italics. The region homologous to a noncoding 3' portion of Rab7 is double underlined.

- Figure 2. Map of the human endogenous retroviral region HERV-7q. The upper part of the figure corresponds to an anonymous region of the human genome situated on the long arm of chromosome 7. The repeat domains (1), *gag* (2), *pol* (3) and *env* (4) of HERV-7q can be identified. The C-terminal *env* region (4.3) is prolonged upstream in the form of a long open reading frame (4.2). The domain 4.1 corresponds to the N-terminal region of the *env* domain.

- Figure 3. Comparison of the repeat nucleic sequences situated at the boundaries of HERV-7q. The 5' (top) and 3' (bottom) repeat nucleic regions are compared and the identical bases are indicated by two dots.

- Figure 4. Deduced sequence having an open reading frame in the env-type domain of HERV-7q according to the longest open reading frame rule.

5 - Figure 5. Sequences around the CKS-17 domain identified in various deduced env domains of the HERV-7q family and comparison with reference CKS-17 motifs.

10 1) HE2 - 2) HERV-7q - 3) GenBank accession No.: M85205 - 4) HE7 - 5) HE9 - 6) CKS-17; the peptide motif endowed with immunomodulatory properties is underlined - 7) gp20 of retrovirus type D (SRV-Pc).

15 - Figure 6. Possible deduced sequence of the gag-type domain identified in HERV-7q established according to the longest open reading frame rule. X and / correspond to a non-sense codon and to a reading frame shift, respectively. The underlined sequence corresponds to the beginning of the pol domain.

20 - Figure 7. Comparison of the nucleic regions covering the gag region of HERV-7q (top) and HERV-TcR (bottom) and their flanking regions. The identical bases are specified by two dots.

25 - Figure 8. Example of nucleic alignments of the env-type domain of HERV-7q with similar env-type domains present in human endogenous retroviral sequences of the same family. The non-sense codons are underlined: 1) HERV-7q - 2) HE2 03) HE3 - 04) HE4.

30 - Figure 9. Nucleic alignments between the gag domain of HERV-7q and the corresponding domains belonging to the same family. Comparison with fragments of gag domains isolated from infectious retroviral agents. Sequences of infectious retroviral origin: EMBL database accession No.: 1) A60168 - 2) A60201 - 3) A60200 - 4) A60171. Human endogenous retroviral sequences: 5) HERV-7q - 6) HG11 - 7) HG3. The figures indicated in the endogenous sequences correspond to the number of nucleotides inserted in order to optimize the alignment with the gag-type sequences identified in retroviruses of infectious origin.

- Figure 10. Alignment of a deduced *gag* protein motif (top) belonging to an infectious retrovirus (EMBL accession No.: A60200) with the deduced *gag* protein motif (bottom) identified in HERV-7q. The non-sense codons are in bold and underlined. The identical amino acids are specified by 2 dashes. One dash indicates a deletion or a homologous amino acid.

- Figure 11. Alignment of an *env* motif (top) belonging to an infectious retrovirus (EMBL accession No.: A60170) with the *env* motif (bottom) identified in HERV-7q. The homologous nucleotides are specified by two dots and the deletions by a dash.

- Figure 12. Comparison between the *env* domain of HERV-7q (top) and the *env* domain of HERV-9 (bottom). The 66% homology is limited to the 3' region of the *env* domain of HERV-7q and HERV-9, respectively between nucleotides 8976 nt and 9500 nt of HERV-7q and nucleotides 2898 nt and 3465 nt of HERV-9 (GenBank accession No.: X57147). Numerous insertions/deletions are also observed.

- Figure 13. Comparison between the *env*-type domains from HERV-7q and from an infection exogenous retroviral sequence (EMBL accession No. A60170).

It should be clearly understood, however, that these examples are given solely by way of illustration of the subject of the invention and do not in any manner constitute a limitation thereto.

EXAMPLE 1: Detection, by gene amplification, of a nucleic sequence belonging to a domain of the *gag* or *env* type according to the invention, in a genomic DNA sample of human or mammalian origin

The gene amplification is carried out using genomic DNA isolated from blood. An anticoagulant treatment is carried out with 1 ml of a citrate solution (per liter: 4.8 g of citric acid, 13.2 g of sodium citrate, 14.7 g of glucose) per 6 ml of fresh blood. After centrifugation of 20 ml of blood for 15 min at 130 000 g, the supernatant is removed and the

fraction enriched with white blood cells is transferred into a new tube and then recentrifuged under the same conditions as above. The fraction enriched with white blood cells is resuspended in an extraction buffer (10 nM Tris-HCl, 0.1 M EDTA, 20 µg/ml of pancreatic RNase treated so as to eliminate the DNases, 0.5% SDS, pH 8.0), and then incubated for 1 hour at 37°C. Proteinase K is added at a final concentration of 100 µg/ml. The suspension of lyzed cells is incubated at 50°C for 3 hours, with occasional stirring, and then treated with an equal volume of phenol equilibrated with 0.5 M Tris-HCl, pH 8.0. The emulsion formed is placed on a wheel for one hour and then centrifuged at 5 000 g for 15 min at room temperature. The aqueous solution is treated and deproteinized by a triple phenol extraction in order to obtain a level of purification corresponding to an absorbance A260/A280 final ratio greater than 1.75. The aqueous fraction is precipitated with 0.2 vol. of 10 M sodium acetate and 2 vol. of ethanol. The DNA is then either collected with the tip of a bent Pasteur pipette, or centrifuged at 5 000 g for 5 min at room temperature. The DNA or the DNA pellet is washed twice with 70% ethanol and then taken up in 1 ml of TE, pH 8.0 so as to be eluted, with gentle stirring, for 12 to 24 hours.

Oligonucleotides specific for the endogenous sequences described according to the invention are chosen in order to amplify the *gag* or *env* region of the endogenous retroviral regions described according to the invention. The genomic DNA studied is obtained from patients having pathological conditions such as multiple sclerosis and from individuals reputed to be healthy.

The thermostable DNA polymerases used were chosen for their high accuracy during the amplification process, such as Vent DNA polymerase (Biolabs) and the like, and are used according to the conditions recommended by the supplier.

The amplification strategy uses, depending on the case, a simple PCR, or a nested or seminested PCR.

Oligonucleotides used to amplify the *gag* region:

- 5 - primer G1F, sense, located in the region upstream of the *gag* domain of *HERV-7q* (SEQ ID NO: 30),
- primer G1R, antisense, located in the 3' terminal region of the *gag* domain (SEQ ID NO: 31).

10 The fragment of 1505 nt amplified by the pair G1F-G1R; 1505 nt is used to generate the probes capable of hybridizing the various PCR amplification products.

- primer G2F, sense nested (SEQ ID NO: 32),
- primer G2R, antisense nested (SEQ ID NO: 33),
- primer G4F, sense nested (SEQ ID NO: 34),
- 15 - primer G3F, sense nested (SEQ ID NO: 35),
- primer G4R, antisense nested (SEQ ID NO: 36),
- primer G5R, antisense nested (SEQ ID NO: 37).

Oligonucleotides used to amplify the *env* region of *HERV-7q*:

- 20 - primer E1F, sense (SEQ ID NO: 38),
- primer E1R, antisense (SEQ ID NO: 39).

25 The fragment of 2529 nt amplified by the pair of primers E1F-E1R is used to generate the probes capable of hybridizing the various PCR amplification products.

- primer E2F, sense (SEQ ID NO: 40),
- primer E2R, antisense (SEQ ID NO: 41),
- primer E3F, sense (SEQ ID NO: 42),
- primer E3R, antisense (SEQ ID NO: 43),
- 30 - primer E4F, sense (SEQ ID NO: 44),
- primer E4R, antisense (SEQ ID NO: 45),
- primer E5F, sense (SEQ ID NO: 46),
- primer E6F, sense (SEQ ID NO: 47),
- primer E5R (SEQ ID NO: 48),
- 35 - primer ExF (SEQ ID NO: 49),
- primer ExR (SEQ ID NO: 50).

The PCR is carried out using 50 to 200 ng of genomic DNA. The PCR conditions are those recommended

by the supplier. The amplification cycle conditions are carried out in 50 µl: denaturation of 94°C for 1 min, hybridization of 70°C for 1 min, and extension at 72°C for 1 to 2 min, depending on the amplified fragments. 5 After 35 cycles, a terminal reaction is carried out at 72°C for 10 min. Automated sequencing of the amplified samples is carried out with the aid of an Applied Biosystems type ABI 377 sequencer or another comparable model, according to the protocols provided by the 10 manufacturer.

In the case of a nested or seminested PCR, the same experimental conditions are used, the only difference being that the genomic DNA sequence is replaced with 5 to 10 µl of the amplification product 15 derived from the first PCR.

Two independent amplifications are carried out using the same sample. A control reaction is carried out by replacing the DNA sample with water in order to detect possible contaminants.

20 **EXAMPLE 2: Detection, by gene amplification, of a nucleic sequence according to the invention in a biological sample of genomic DNA collected from patients having an existing candidate pathological condition or suspected of having this pathological**
25 **condition**

The amplification protocol is the same as in Example 1, apart from the origin of the sample which is obtained from patients having a candidate pathological condition. A genomic DNA sample reputed to be normal is 30 systematically integrated into the set of amplified pathological samples and then analyzed.

The PCR products are separated on a 1.5% agarose gel and then transferred in the presence of 0.4 N sodium hydroxide on a charged nylon membrane. 35 Hybridization is carried out with a specific probe corresponding to the PCR fragments amplified either with the pair G1F-G1R or the pair E1F-E1R. The probe is labeled by incorporating dUTP-digoxigenin according to

the supplier's protocol (Boehringer Mannheim). The hybridization is carried out in a hybridization buffer (5XSSC, 50% formamide, 0.1% lauroylsarcosine, 0.02% SDS, 2% blocking reagent Boehringer) overnight at 42°C. The Southern is washed for twice 5 min at room temperature in a 2XSSC solution containing 0.1% SDS. Next, a high stringency wash is carried out twice for 15 min at 55°C in a 0.1XSSC solution containing 0.1% SDS. The hybridization is visualized according to the supplier's protocol (Boehringer Mannheim), in the presence of a chemiluminescent substrate for alkaline phosphatase, of the CSPD or CDP-STAR type. The filter is visualized after a 15 min exposure at 60°C.

SSCP (*single strand conformation polymorphism*) analysis makes it possible to detect discrete modifications of the sequence of the fragments amplified by PCR. The PCR is carried out in the presence of dCTP labeled with ^{32}P . The sample to be analyzed is denatured at 95°C for 10 min in the presence of loading buffer, and then immediately loaded onto a 10% polyacrylamide gel containing 7.5% glycerol. The migration is carried out at 4°C at 8-10 W. The gel is dried and then autoradiographed.

The PCR fragments likely to exhibit an alteration of their nucleotide sequence are sequenced according to Example 1.

Hybridization with the aid of a specific oligonucleotide (17 mers to 20 mers) corresponding to the modified nucleotide region makes it possible to identify the samples having an identical modification (ASO method). Briefly, the southern is hybridized with an oligonucleotide which is distally labeled either with ^{32}P , or in the presence of digoxigenin (according to the Boehringer Mannheim protocol) and then washed under stringent conditions at 65°C in a 6XSSC solution containing 0.05% sodium pyrophosphate.

EXAMPLE 3: Detection of a protein according to the invention in a biological sample

- Preparation of a purified protein fraction of cerebrospinal fluid from patients suffering from MS

After a treatment at 56°C for 30 min and removal of the immunoglobulins on a G HiTrap protein column (Pharmacia), the equivalent of 10 ml of CSF is deposited on a DEAE Sepharose CL-6B column (Pharmacia). The elution is carried out in 20 mM Tris-HCl, pH 8.8, and a gradient from 0 to 0.4 M NaCl, and then the fraction is dialyzed twice against a phosphate-NaCl buffer (PBS). After concentration on Ultrafree-MC (Millipore), the fraction is deposited on a Superose 12 column (FPLC Pharmacia) and eluted in the presence of PBS. After separation by polyacrylamide-SDS gel electrophoresis and electrotransfer onto an Immobilon-P membrane (Millipore), the protein bands are subjected to controlled trypsin hydrolysis.

- Analysis of the protein fraction by mass spectrometry

The peptides digested in the presence of trypsin are analyzed by the MALDI-TOF method, which allows the analysis of peptides present in a mixture (COTTRELL J.S., Pept. Res., 1997, 7, 115-124). The peptides characterized according to their mass are compared with the proteins and with the associated proteins according to the invention.

EXAMPLE 4: Detection of specific antibodies to the env domain of HERV-7q

The identification of a long open reading frame in the env sequence of HERV-7q made it possible to determine a deduced protein sequence SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29 of a region of the said gene referenced by SEQ ID NO: 22.

The protein sequences deduced from the sequences ID NO: 23, 25, 27, 28, 29 are positioned as follows with respect to Figure 1 or the sequence ID NO: 3:

SEQ ID NO: 23 beginning of the coding sequence:
position 7874, end of the coding sequence 1st nonsense
codon (position 9493)

5 SEQ ID NO: beginning of the coding sequence:
position 7874, end of the coding sequence 1st nonsense
codon (position 9493) (reading frame 1)

SEQ ID NO: 27 beginning of the coding sequence:
position 6970, end of the coding sequence 1st nonsense
codon (position 9493) (reading frame 1)

10 SEQ ID NO: 28 beginning of the coding sequence:
position 6971, the end of the reading frame is shifted
depending on the case by 1, 2 or 3 codons

SEQ ID NO: 29 beginning of the coding sequence:
position 6972, the end of the reading frame is shifted
15 depending on the case by 1, 2 or 3 codons

Various peptides corresponding to all or part
of SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID
NO: 28, SEQ ID NO: 29 were synthesized by genetic
engineering in order to test their antigenic
20 specificity toward sera or tissues from patients
suffering from MS, for example. Briefly, all or part of
the env region of HERV-7q is subcloned into the vectors
pQE30, 31 and 32. The vectors pQE30, 31 and 32 contain,
in 5' of the multiple cloning site, the consensus
25 sequences for transcription (the strong T5
bacteriophage promoter, 2 operators of the lactose
operon) and translation (one synthetic ribosome binding
site). Likewise, pQE30, 31 and 32 possess, in 3', the
phage 1 transcription terminator as well as a Stop
30 codon for translation. The expression of the protein is
carried out after transformation in *E. coli* M15. The
plasmid pQE30, 31 and 32 possess, upstream of the
multiple cloning site, the coding sequence for a
succession of 6 histidines having affinity for nickel
35 ions. This stretch allows the purification of the
expressed chimeric protein by adsorption on a resin
consisting of a chelating ligand, nitrotriacetic acid

(NTA), charged with 4 nickel ions (NI-NTA resin, Qiagen).

The transformation is carried out by electroporation or treatment with calcium chloride. For example, an *E. coli* M15 colony is incubated in 100 ml of LB medium containing 250 µg of kanamycin, with stirring at 37°C until an OD⁶⁰⁰ of 0.5 is obtained. After centrifugation for 5 minutes at 2000 g at 4°C, the bacterial pellet is taken up in 30 ml of TFB1 solution (100 mM rubidium chloride, 50 mM manganese chloride, 30 mM potassium acetate, 10 mM CaCl₂, 15% glycerol, pH 5.8), at 4°C for 90 minutes. After a centrifugation of 5 minutes at 2000 g at 4°C, the bacterial pellet is taken up in 4 ml of TFB2 solution (10 mM rubidium chloride, 10 mM MOPS, 75 mM CaCl₂, 15% glycerol, pH 8). The cells may be kept at -70°C in aliquots of 500 µl. 20 µl of the ligation and 125 µl of competent cells are mixed and placed on ice for 20 minutes. After a heat shock of 42°C for 90 seconds, the cells are stirred for 90 minutes at 37°C in 500 µl of Psi-broth medium (LB medium supplemented with 4 mM MgSO₄, 10 mM potassium chloride). The transformed cells are plated on LB-agar dishes supplemented with 25 µg/ml of kanamycin and 100 µg/ml of ampicillin, and the dishes are incubated overnight at 37°C.

The potentially recombinant clones are subcultured in an orderly manner on a nylon filter deposited on an LB-agar dish supplemented with 25 µg/ml of kanamycin and 100 µg/ml of ampicillin. After one night at 37°C, the recombinant clones are located by hybridization of the plasmid DNA with the nucleotide probe amplified by PCR with the pair of primers according to SEQ ID NO: 38 and SEQ ID NO: 39.

An independent colony containing the insert is inoculated at 20 ml of LB medium supplemented with 25 µg/ml of kanamycin and 100 µg/ml of ampicillin. After one night at 37°C, with stirring, 500 µl of the same medium are incubated at 1/50 with this preculture

until an OD⁶⁰⁰ of 0.8 is obtained, and then 1 to 2 mM final of IPTG is added. After 5 hours, the cells are centrifuged for 20 minutes at 4 000 g.

5 A portion of the cellular pellet is taken up
in 5 ml of sonification buffer (50 mM of sodium
phosphate, pH 7.8, 300 mM NaCl) and then placed on ice.
After rapid sonification, the cells are centrifuged for
20 minutes at 10 000 g. A portion of the cellular
10 pellet is taken up in 10 ml of a 30 mM Tris/HCl-20%
sucrose solution pH 8. The cells are incubated for 5 to
10 minutes, with stirring, after addition of 1 mM EDTA.
After a centrifugation of 10 minutes at 8 000 g at 4°C,
the pellet is taken up in 10 ml of 5 mM ice cold MgSO₄.
After 10 minutes on the ice, with stirring, the cells
15 are centrifuged for 10 minutes at 8 000 g at 4°C.

The pellet is taken up in 5 ml/g in buffer A
(6 M GuHCl (guanidine hydrochloride), 0.1 M sodium
phosphate, 0.01 M Tris/HCl, pH 8), 1 hour at room
temperature. The lysate is centrifuged for 15 minutes
20 at 10 000 g at 4°C, and the supernatant is supplemented
with 8 ml of Ni-NTA resin, pre-equilibrated in
buffer A. After 45 minutes at room temperature, the
resin is poured into a column, washed with 10 times the
column volume with buffer A and then with 5 times the
25 column volume with buffer B (8 M urea, 0.1 M sodium
phosphate, 0.01 M Tris/HCl, pH 8). The column is washed
with buffer C (8 M urea, 0.1 M sodium phosphate, 0.01 M
Tris/HCl, pH 6.3) until A280 is less than 0.01. The
recombinant protein is eluted with 10 to 20 ml of
30 buffer D (8 M urea, 0.1 M sodium phosphate, 0.01 M
Tris/HCl, pH 5.9) and then with 10 to 20 ml of buffer E
(8 M urea, 0.1 M sodium phosphate, 0.01 M Tris/HCl,
pH 4.5), and then with 20 ml of buffer F (6 M HCl,
0.2 M acetic acid). After SDS-PAGE analysis, the
35 purified fraction(s) containing the chimeric protein
allowed the production of antibodies in rabbits. The
antibodies obtained are tested by Western blotting

after visualization with a secondary antibody coupled to alkaline phosphatase.

Antibodies are obtained in the same manner, using peptides synthesized chemically according to the Merrifield technique (G. Barany and B. Merrifield, 5 Merrifield technique (G. Barany and B. Merrifield, 1980, in *The peptides*, 2, 1-284, E. Gross and J. Meienhofer, Academic Press, New York).

The specific antibodies obtained are used for detection of the serum or tissue expression of all or part of the endogenous retroviral sequences according to the invention, in normal and pathological cases. 10

The proteins of serum or tissue origin are separated on acrylamide-SDS gel and then transferred onto a nitrocellulose filter with the aid of a Novablot 2117-2250 apparatus (LKB). The transfer is carried out on a Hybond C-extra sheet (Amersham) using a 100 mM CAPS buffer pH 11, methanol, water (V/V/V: 1/1/8) containing 1 mM CaCl_2 . After a transfer of 1 hour at 0.8 mA/cm², the sheet is saturated for 1 hour at room 20 temperature in PBS-0.5% gelatin. The sheet is brought into contact with the specific antibody at the concentration of 1/1 000 in PBS-0.25% gelatin. After 2 hours, the filter is washed 3 times 15 minutes in PBS-0.1% Tween-20, and then the filter is incubated for 25 30 minutes in the presence of a secondary antibody coupled to alkaline phosphatase (Promega), diluted 1/7 500 in PBS-0.25% gelatin. After three washes in PBS-0.1% Tween-20, the filter is equilibrated in a buffer (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM 30 MgCl_2). The visualization is carried out in the presence of 45 μl of NBT at 75 mg/ml and 35 μl of BCIP at 50 mg/ml, per 10 ml of alkaline phosphatase buffer.

The chimeric proteins obtained by genetic engineering are also used for tests of biological activity, such as for example the test for biological 35 activity of the CKS-17-type peptide identified in the env domain of HERV-7q (Figure 5).

EXAMPLE 5: Production of ribonucleic probes encoding the env sequences of HERV-7q

The PCR fragments obtained are subcloned into the plasmid PGEM 4Z (Promega) which possesses on either side of its multiple cloning site, promoter sequences for the SP6 and T7 RNA polymerases.

The method of competence used is electroporation. The plasmid and the PCR fragment are hybridized in a ratio of 50 ng of vector (SmaI cleavage) to 100 ng of PCR fragment (made blunt ended by treatment with the Klenow fragment of DNA polymerase). The incubation takes place overnight at 22°C in ligation buffer (66 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 1 mM dithioerythritol, 1 mM ATP) in the presence of 1 u of T4 DNA ligase and is then stopped by denaturation for 10 minutes at 65°C. In parallel, the *E. coli* JM 105 strain is inoculated overnight at 37°C in LB medium. This preculture is diluted 1/500 and placed at 37°C until an OD⁶⁰⁰ equal to 1 is obtained. For the remainder of the procedure, the cells will always be stored at cold temperature. After centrifugation for 5 minutes at 3 500 g at 4°C, the cellular pellet is resuspended in 1/4 vol. of ultra-pure ice-cold water. This step is repeated 5 to 6 times. The pellet is then resuspended in 1/4 000 vol. of water; 10% of sterile glycerol is added, allowing preservation of the electrocompetent cells, in aliquots of 10 µl at 20°C. 1 µl of the ligation is added to 50 µl of electrocompetent cells; the mixture is subjected to an electrical discharge of 12.5 kV/cm, applied for 5.8 ms. The cells are rapidly resuspended in the SOC medium, incubated for 1 hour at 37°C and then plated in the presence of 2% X-Gal in dimethylformamide, and 10 mM IPTG, on an LB-agar dish supplemented with ampicillin (100 µg/ml). After one night at 37°C, the potentially recombinant white clones are subcultured in an orderly manner on an LB/ampicillin dish and in parallel on a nylon filter

deposited on an LB/ampicillin dish. These two dishes are incubated overnight at 37°C. The recombinant clones are then located by hybridization with a nucleic probe amplified by PCR with the pair of primers according to
5 SEQ ID NO: 38 and SEQ ID NO: 39 and labeled with digoxigenin.

The recombinant clones are cultured in 50 ml of LB/ampicillin medium (100 µg/ml), with stirring, overnight at 37°C. After centrifugation at 3 500 g for
10 15 minutes at 4°C, the bacterial pellet is taken up in 4 ml of P1 buffer (50 mM Tris-HCl, 10 mM EDTA, 400 µg/ml RNase A, pH 8) and 4 ml of P2 buffer (200 mM NaOH, 1% SDS). The medium is incubated at room temperature for 5 minutes. After addition of 4 ml of
15 P3 buffer (2.55 M potassium acetate, pH 4.8), the mixture is centrifuged at 12 000 g for 30 minutes at 4°C. This supernatant is applied to a Qiagen type 100 column, pre-equilibrated with 2 ml of QBT buffer (750 mM NaCl, 50 mM MOPS, 15% ethanol, pH 7), the
20 column is washed with twice 4 ml of QC buffer (1 M NaCl, 50 mM MOPS, 15% ethanol, pH 7) and the DNA is eluted with 2 ml of QF buffer (1.2 M NaCl, 50 mM MPOS, 15% ethanol, pH 8). The DNA is precipitated with 0.8 vol. of isopropanol and centrifuged at 12 000 g at
25 4°C for 30 minutes. The pellet is washed with 70% ice-cold ethanol and then the plasmid DNA is taken up in twice 150 µl of TE buffer.

The ribonucleic probes are used as specific probes, in particular for the detection of the
30 transcripts expressed by the endogenous retroviral sequences according to the invention.

Bibliographic references:

- Benit L. et al., 1997. Cloning of a new murine endogenous retrovirus MuERV-L, with strong similarity
35 of the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. J. Virol. 71, 5652-5657.

- Coffin J.M. 1985. Endogenous retrovirus, In: "RNA tumor viruses" (Weiss R.A., Varmus H.E., Teich N.M., and Coffin J.M. eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- 5 - Conrad B., Weissmahr R.N., Boni J., Arcari R., Schupbach J., and Mach B. 1997. A human endogenous retroviral superantigen as candidate autoimmunogene in type 1 diabetes. *Cell* 90, 303-313.
- Covey S.N. 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus, *Nucleic Acids Res.* 14, 623-633.
- 10 - Hertig C., Coupar B.E., Gould A.R., and Boyle D.B. 1997. Field and vaccine strains of fowlpox virus carry integrated sequences from the avian retrovirus, reticuloendotheliosis virus. *Virology* 235, 367-376.
- 15 - Hohenadl C., Leib-Mösch C., Hehlemann R., and Erfle Y. 1996. Biological significance of human endogenous retroviral sequences. *J. Acqui. Imm. Def. Synd. Hum. Retrovir.* 13, S268-S273.
- 20 - Kulkoski J.K., Jones S., Katz R.A., Mack J.P.G., and Skalka A.M. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* 12, 2331-2338.
- La Mantia G. et al., *N.A.R.*, 1991, 19, 7, 1513-1520
- Patience C., Wilkinson D.A., and Weiss R.A. 1997. Our retroviral heritage. *Trends Genet.* 13, 116-120.
- 30 - Pearson W.R. 1994. Using the FASTA program to search protein and DNA sequence databases. *Methods Mol. Biol.* 24, 307-331.
- Perron H., Garson J.A., Bedin F., Beseme F., Paranhos-Baccala G., Komurian-Pradel F., Mallet F., Tuke P.W.,
- 35 Voisset C., Blond J.L., Lalande B., Seigneurin J.M., Mandrand B. and the Collaborative Research Group on Multiple Sclerosis. 1997. Molecular identification of a novel retrovirus repeatedly isolated from patients

- with multiple sclerosis. Proc. Natl. Acad. Sci. USA 94, 7583-7588.
- Tönjes R.R. et al., J. AIDS and Hum. Retrovirol. 1996, 13. S261-S267.
- 5 - Vitelli R., Chiarillo M., Lattero D., Bruni C.B., and Bucci C., 1996. Molecular cloning and expression analysis of the human Rab7 GTP-ase complementary deoxyribonucleic acid. Biochem. Biophys. Res. Commun. 229, 887-890.
- 10 - Weber L.T., Miller M., Jaskolski M., Leis J., Skalka M., and Wlodawer A., 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. Science 243, 928-931.
- Wilkinson D., Mager D.L., and Leong J.A.C. 1994.
- 15 Endogenous human retroviruses. In: "The Retroviridae" (Levy J.A. ed). Plenum Press New York., Vol. 3, 465-535.
- Xiong Y., and Eickbush, T. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353-3362.
- 20

As is evident from the above, the invention is not at all limited to its embodiments, implementations and applications which have just been described more explicitly; it embraces on the contrary all the

25 variants which may occur to a specialist in this field, without departing from the framework or scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: INSTITUT NATIONAL DE LA RECHERCHE MEDICALE -
INSERM
(B) ROAD: 101 RUE DE TOLBIAC
(C) TOWN: PARIS
(E) COUNTRY: FRANCE
(F) POSTAL CODE: 75654 CEDEX

(ii) INVENTION TITLE: NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE
FAMILY WITH HUMAN ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES.

(iii) NUMBER OF SEQUENCES: 51

(iv) COMPUTER READABLE FORM

(A) SUPPORT TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (OEB)

(2) INFORMATION FOR SEQ ID NO: 1: env

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2599 base pairs
(B) TYPE: nucleotide
(C) STRANDS NUMBER: single
(D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE : DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATCCCCTGCC TTAATCGCCA AGCTCCTTCA GGAGAACAAA GAACAGGCCA TTACCCTGGA	60
GAAGACTGGC AACTGATTTT ACCACAAGC CCAAACCTCA GGGATTTCAG TATCTACTAG	120
TCTGGGTAGA TACTTTCACG GGTGGGCAG AGGCCTTCCC CTGTAGGACA GAAAAGGCC	180
AAGAGGTAAT AAAGGCACTA GTTCATGAAA TAATTCCCAG ATTCGGACTT CCCCAGGCT	240
TACAGAGTGA CAATAGCCCT GCTTTCAGG CCACAGTAAC CCAGGGAGTA TCCCAGGCGT	300
TAGGTATACG ATATCACTTA CACTGCGCCT GAAGGCCACA GTCCTCAGGG AAGGTCGAGA	360
AAATGAATGA AACACTCAA GGACATCTAA AAAAGCAAAC CCAGGAAACC CACCTCACAT	420
GGCCTGCTCT GTTGCCTATA GCCTTAAAAA GAATCTGCAA CTTTCCCCAA AAAGCAGGAC	480
TTAGCCCATA CGAAATGCTG TATGGAAGGC CCTTCATAAC CAATGACCTT GTGCTTGACC	540
CAAGACAGCC AACTTAGTTG CAGACATCAC CTCCTTAGCC AAATATCAAC AAGTTCTTAA	600
AACATTACAA GGAACCTATC CCTGAGAAGA GGGAAAAGAA CTATTCCACC CTTGTGACAT	660

GGTATTAGTC	AAGTCCCTTC	CCTCTAATTC	CCCATCCCTA	GATACATCCT	GGGAAGGACC	720
CTACCCAGTC	ATTTTATCTA	CCCCAACTGC	GGTTAAAGTG	GCTGGAGTGG	AGTCTTGGAT	780
ACATCACACT	TGAGTCAAAT	CCTGGATACT	GCCAAAGGAA	CCTGAAAATC	CAGGAGACAA	840
CGCTAGCTAT	TCCTGTGAAC	CTCTAGAGGA	TTTGCGCCTG	CTCTTCAAAC	AACAACCAGG	900
AGGAAAGTAA	CTAAAATCAT	AAATCCCCAT	GGCCCTCCCT	TATCATATTT	TTCTCTTTAC	960
TGTTCTTTTA	CCCTCTTTCA	CTCTCACTGC	ACCCCCTCCA	TGCCGCTGTA	TGACCAGTAG	1020
CTCCCCTTAC	CAAGAGTTTC	TATGGAGAAT	GCAGCGTCCC	GGAAATATTG	ATGCCCCATC	1080
GTATAGGAGT	CTTTCTAAGG	GAACCCCCAC	CTTCACTGCC	CACACCCATA	TGCCCCGCAA	1140
CTGCTATCAC	TCTGCCACTC	TTTGCATGCA	TGCAAATACT	CATTATTGGA	CAGGAAAAAT	1200
GATTAATCCT	AGTTGTCCTG	GAGGACTTGG	AGTCACTGTC	TGTTGGACTT	ACTTCACCCA	1260
AACTGGTATG	TCTGATGGGG	GTGGAGTTCA	AGATCAGGCA	AGAGAAAAAC	ATGTAAAGA	1320
AGTAATCTCC	CAACTCACCC	GGGTACATGG	CACCTCTAGC	CCCTACAAAG	GACTAGATCT	1380
CTCAAAACTA	CATGAAACCC	TCCGTACCCA	TACTCGCCTG	GTAAGCCTAT	TTAATACCAC	1440
CCTCACTGGG	CTCCATGAGG	TCTCGGCCCA	AAACCCTACT	AACTGTTGGA	TATGCCTCCC	1500
CCTGAACTTC	AGGCCATATG	TTTCAATCCC	TGTACCTGAA	CAATGGAACA	ACTTCAGCAC	1560
AGAAATAAAC	ACCACTTCCG	TTTTAGTAGG	ACCTCTTGTT	TCCAATCTGG	AAATAACCCA	1620
TACCTCAAAC	CTCACCTGTG	TAAAATTTAG	CAATACTACA	TACACAACCA	ACTCCCAATG	1680
CATCAGGTGG	GTAACCTCTC	CCACACAAAT	AGTCTGCCTA	CCCTCAGGAA	TATTTTTTGT	1740
CTGTGGTACC	TCAGCCTATC	GTTGTTTGAA	TGGCTCTTCA	GAATCTATGT	GCTTCCTCTC	1800
ATTCTTAGTG	CCCCCTATGA	CCATCTACAC	TGAACAAGAT	TTATACAGTT	ATGTCATATC	1860
TAAGCCCCGC	AACAAAAGAG	TACCCATTCT	TCCTTTTGTT	ATAGGAGCAG	GAGTGCTAGG	1920
TGCACTAGGT	ACTGGCATTG	GCGGTATCAC	AACCTCTACT	CAGTTCTACT	ACAAACTATC	1980
TCAAGAACTA	AATGGGGACA	TGGAACGGGT	CGCCGACTCC	CTGGTCACCT	TGCAAGATCA	2040
ACTTAACTCC	CTAGCAGCAG	TAGTCCTTCA	AAATCGAAGA	GCTTTAGACT	TGCTAACCGC	2100
TGAAAGAGGG	GGAACCTGTT	TATTTTTTAGG	GGAAGAATGC	TGTTATTATG	TTAATCAATC	2160
CGGAATCGTC	ACTGAGAAAG	TTAAAGAAAT	TCGAGATCGA	ATACAACGTA	GAGCAGAGGA	2220
GCTTCGAAAC	ACTGGACCCT	GGGGCCTCCT	CAGCCAATGG	ATGCCCTGGA	TTCTCCCCTT	2280
CTTAGGACCT	CTAGCAGCTA	TAATATTGCT	ACTCCTCTTT	GGACCCTGTA	TCTTTAACCT	2340
CCTTGTTAAC	TTTGTCTCTT	CCAGAATCGA	AGCTGTAAAA	CTACAAATGG	AGCCCAAGAT	2400
GCAGTCCAAG	ACTAAGATCT	ACCGCAGACC	CCTGGACCGG	CCTGCTAGCC	CACGATCTGA	2460
TGTTAATGAC	ATCAAAGGCA	CCCCTCCTGA	GGAAATCTCA	GCTGCACAAC	CTCTACTACG	2520

CCCCAATTCA GCAGGAAGCA GTTAGAGCGG TCTCGGCCAA CCTCCCCAAC AGCACTTAGG 2580
TTTTCCTGTT GAGATGGGG 2599

(2) INFORMATIONS FOR SEQ ID NO: 2: gag

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1326 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE : DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCCGCCTGGC ACTCCTGAGG GAAGTATAAA TTATAACACC ATCTTACAGC TAGACCTCTT 60
TTGTAGAAAA GGCAAATGGA GTGAAGTGCC ATAAGTACAA ACTTTCCTTTT CATTAAGAGA 120
CAACTCACAA TTATGTAAAA AGTGTGATTT ATGCCCTACA GGAAGCCTTC AGAGTCTACC 180
TCCCTATCCC AGCATCCCCG ACTCCTTCCC CAACTAATAA GGACCCCCCT TCAACCCAAA 240
TGGTCCAAAA GGAGATAGAC AAAAGGGTAA ACAGTGAACC AAAGAGTGCC AATATTCCCC 300
AATTATGACC CCTCCAAGCA GTGGGAGGAA GAGAATTCGG CCCAGCCAGA GTGCATGTGC 360
CTTTTTCTCT CCCAGACTTA AAGCAAATAA AAACAGACTT AGGTAAATTC TCAGATAACC 420
CTGATGGCTA TATTGATGTT TTACAAGGGT TAGGACAATT CTTTGATCTG ACATGGAGAG 480
ATATAATGTC ACTGCTAAAT CAGACACTAA CCCCAAATGA GAGAAGTGCC ACCATAACTG 540
CAGCCTGAGA GTTTGGCGAT CTCTGGTATC TCAGTCAGGT CAATGATAGG ATGACAACAG 600
AGGAAAGAGA ATGATTCCCC ACAGGCCAGC AGGCAGTTCC CAGTCTAGAC CCTCATTGGG 660
ACACAGAATC AGAACATGGA GATTGGTGCT GCAGACATTT GCTAACTTGT GTGCTAGAAG 720
GACTAAGGAA AACTAGGAAG AAGTCTATGA ATTACTCAAT GATGTCCACC ATAACACAGG 780
GAAGGGAAGA AAATCCTACT GCCTTTCTGG AGAGACTAAG GGAGGCATTG AGGAAGCGTG 840
CCTCTCTGTC ACCTGACTCT TCTGAAGGCC AACTAATCTT AAAGCGTAAG TTTATCACTC 900
AGTCAGCTGC AGACATTAGA AAAAACTTC AAAAGTCTGC CGTAGGCCCG GAGCAAAACT 960
TAGAAACCTT ATTGAACTTG GCAACCTCGG TTTTTTATAA TAGAGATCAG GAGGAGCAGG 1020
CGGAACAGGA CAAACGGGAT TAAAAAAAAG GCCACCGCTT TAGTCATGAC CCTCAGGCAA 1080
GTGGACTTTG GAGGCTCTGG AAAAGGGAAA AGCTGGGCAA ATTGAATGCC TAATAGGGCT 1140
TGCTTCCAGT GCGGTCTACA AGGACACTTT AAAAAAGATT GTCCAAGTAG AAGTAAGCCG 1200
CCCCCTCGTC CATGCCCTT ATTTCAAGGG AATCACTGGA AGGCCCACTG CCCCAGGGGA 1260

CAAAGGTCCT CTGAGTCAGA AGCCACTAAC CAGATGATCC AGCAGCAGGA CTGAGGGTGC 1320
CTGGGG 1326

(2) INFORMATIONS FOR SEQ ID NO: 3: HERV-7q

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10499 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE : DNA (genomic)

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 3:

CCCTGGGGCG GGCTTCCTTT CTGGGATGAG GGCAAAACGC CTGGAGATAC AGCAATTATC 60
TTGCAACTGA GAGACAGGAC TAGCTGGATT TCCTAGGCCG ACTAAGAATC CCTAAGCCTA 120
GCTGGGAAGG TGACCACGTC CACCTTTAAA CACGGGGCTT GCAACTTAGC TCACACCTGA 180
CCAATCAGAG AGCTCACTAA AATGCTAATT AGGCAAAGAC AGGAGGTAAA GAAATAGCCA 240
ATCATCTATT GCCTGAGAGC ACAGCAGGAG GGACAACAAT CGGGATATAA ACCCAGGCAT 300
TCGAGCTGGC AACAGCAGCC CCCCTTTGGG TCCCTTCCCT TTGTATGGGA GCTGTTTTCA 360
TGCTATTTCA CTCTATTAAA TCTTGCAACT GCACTCTTCT GTCCATGTT TCTTACGGCT 420
CGAGCTGAGC TTTTGCTCAC CGTCCACCAC TGCTGTTTGC CACCACCGCA GACCTGCCGC 480
TGA TCCCAT CCCTCTGGAT CCTGCAGGGT GTCCGCTGTG CTCCTGATCC AGCGAGGCGC 540
CCATTGCCGC TCCCAATTGG GCTAAAGGCT TGCCATTGTT CCTGCACGGC TAAGTGCCCTG 600
GGTTTGTCT AATTGAGCTG AACACTAGTC ACTGGGTTCC ATGGTTCTCT TCTGTGACCC 660
ACGGCTTCTA ATAGAACTAT AACACTTACC ACATGGCCCA AGATTCCATT CTTTGAATC 720
CGTGAGGCCA AGAACTCCAG GTCAGAGAAT ACGAGGCTTG CCACCATCTT GGAAGCGGCC 780
TGCTACCATC TTGGAAGTGG TTCACCACCA TCTTGGGAGC TCTGTGAGCA AGGACCCCCC 840
GGTAACATTT TGGCAACCAC GAACGGACAT CCAAAGTGGT GAGTAATATT GGACCACTTT 900
CACTTGCTAT TCTGTCCTAT CCTTCCTTAG AATTGGAGGA AAATACCGGG CACTTGTCGG 960
CCAGTTAAAA ACGATTAGTG TGGCCACCGG ACTTAAGACT CAGGTGTGAG GCTATCTGGG 1020
GAAGGGCTTT CTAACAACCC CCAACCCTTC TGGGTTGGGG ACTTGGTTTG CCTCAAGCCA 1080
GCTTCCACTT TCAGTTTTCT TGGGGAAGCC GAGGGCCGAC TAGAGGCAGA AAGCTGTCTG 1140
CCTGAACTCC CGGCAGTAGC CGGTTGAGAT CATGGTGTAG CCAGAAGTCT CAACAGTCGC 1200
CCATGCATGC ACCCCTATCT TTCCTTCTGA CCCATACCTC CTGGGTCCCA ACCACAACCT 1260

TCTTCAAAGT	GTAGCCCCAA	AATTCTCCTT	ACCTCTGAAT	ATACTTCCTC	TGATCCCTGC	1320
CTCCTAGGTA	CTATTGGTTC	AGACTTCCAT	TTCCTCTAGC	AAGTTGTATC	TCCAAAGGGA	1380
TCTAAGGAAG	CTCTGCGCTG	CGTCCTTAGG	CACCTAGGCT	ATAACCCAGG	GAGTCTTATC	1440
CCTGGTGTCC	CTCCCAATTT	AGGCATACAG	CTCTTGACAT	GGGCAGTTAT	GTAGGACCCA	1500
CTCCCCACCA	CCCTTGCCAG	GGCCCCAAGT	TTGTAAATGG	CTGAGGGAAA	AGAGAGACAG	1560
AGGAGAGAGA	GAGAAATGGA	GGAGAAAGAG	AGAGAGACAG	AGAGGAGAGA	GAGACAGTGA	1620
GAGAGACAGA	AGAGAGAGAG	AGACAAAGAG	GAGAGAGAGA	GAGTCAAAGA	GAGAAAGAAA	1680
GAGAAAGAAA	TAGTAAAAAA	CAGTGTGCCC	TATTCCTTTA	AAAGCCAGGG	TAAATTTAAA	1740
ACCTGTACTT	GATAATTGAA	GGTCTTCTCT	GTGACCCTAT	AGCACTCCAA	TCCACTTTGT	1800
GGTCAGTGTA	AATAAGAGCA	TAGGCCGAAA	GCACTGAGGC	CATTGACAAC	CCGTAGCTTC	1860
CCTATCAAAA	ATCCTTAACC	CAGTAACCCG	CAGATGGACC	AAATGCATTC	AGTCGGTAGC	1920
GCAACTGCTT	TGCTAAAAGT	AGAAAAGTAA	CTTTTAGAGG	AAACCTCATT	GTGAGCACAC	1980
CTCACCTGTT	CAGAATTATT	CTAATAAAAA	AAGCAAAAAG	GTAGCTTACT	AACTCAAAAA	2040
TCTTAAAGTA	TGGGGCTATT	CTGTTAGAAA	AAGGTAATGT	AACTCCAACC	ACTGATAATT	2100
CCCTTAACCC	AGCAGATTTT	CTAACGGGAT	TTAAATCTTA	ATTACCATAC	AAAGGTCCGA	2160
CCAGACCTAG	GCGGAACTCC	CTTCAGGACA	GGACGATAGA	TGGTTCCTCC	CAGGTGATTG	2220
AGGAAAAAAA	CCACAATGGG	TATTCAGTAA	TTGATACGGG	GACTCTTGTG	GAAGCAGAGT	2280
TAGAAAAAAT	GCCTAATAAC	TGGTCTCCTC	AAACGTGTGA	GCTGTTTGCA	CTCAGCCAAG	2340
CCTTAAAGTA	CTTACAGAAT	CAAAAGACTA	TCTCAATCCT	GATTCAAAAG	GTTAGCTACA	2400
CCCTCTCTGT	AATGCATTTG	CATAAGAACT	TGTTTATGGG	AATGCATCTT	GATGGGGCAG	2460
CTGGGTGTGT	ATAAAATAGG	AACCCAGCCC	AGCTCTAGGA	CTCACCCCTG	AGCGCAAAGG	2520
CAATGTTGGG	CATGCTGGTA	AAGGACCACT	AGAATCCAGC	AGCCCAGACC	CCTTTCTTTG	2580
TGGTCAAGAA	AGGCGGGAAA	AGGGGTGCAG	GA CTGCTACA	TCGGTAAGCA	TA ACTAATCC	2640
GATAAACAGA	GGTCCATGGG	TGGTTACGCA	CCCTGGAAAAG	GAACTCACCC	CTGAGCACAA	2700
AGGCAATGTT	GGGCACGCTG	GTAAAGGACC	ACTAGAATCC	AGCAGCCTGG	ACCCCTTTCT	2760
TTGTGGTCAA	GAGAGGCAGG	AAAACAGGTG	CAGGACTGCA	ACATCAGTGA	GCATAACTAA	2820
TTCGATAAGC	AGAGGTCCAT	GGGTGGTGAT	GCACCCTGGA	AAGAATAAGC	ATTAGGACCA	2880
TAGAGGACAC	TCCAGGACTA	AAGCTCATCG	GAAAATGACT	AGGGTTGCTG	GCATCCCTAT	2940
GTTCTTTTTT	CAGATGGGAA	ACGTTCCCCG	CAAGACAAAA	ACGCCCCTAA	GACGTATTCT	3000
GGAGAATTGG	GACCAATTTG	ACCCTCAGAC	ACTAAGAAAAG	AAACGACTTA	TATTCTTCTG	3060

CAGTGCCGCC	TGGCACTCCT	GAGGGAAGTA	TAAATTATAA	CACCATCTTA	CAGCTAGACC	3120
TCTTTTGTAG	AAAAGGCAAA	TGGAGTGAAG	TGCCATAAGT	ACAAACTTTC	TTTTCATTA	3180
GAGACAACTC	ACAATTATGT	AAAAAGTGTG	ATTTATGCCC	TACAGGAAGC	CTTCAGAGTC	3240
TACCTCCCTA	TCCCAGCATC	CCCGACTCCT	TCCCCAACTA	ATAAGGACCC	CCCTTCAACC	3300
CAAATGGTCC	AAAAGGAGAT	AGACAAAAGG	GTAAACAGTG	AACCAAAGAG	TGCCAATATT	3360
CCCCAATTAT	GACCCCTCCA	AGCAGTGGGA	GGAAGAGAAT	TCGGCCCAGC	CAGAGTGCAT	3420
GTGCCTTTTT	CTCTCCCAGA	CTTAAAGCAA	ATAAAAAACAG	ACTTAGGTAA	ATTCTCAGAT	3480
AACCCTGATG	GCTATATTGA	TGTTTTACAA	GGGTTAGGAC	AATTCTTTGA	TCTGACATGG	3540
AGAGATATAA	TGTCACTGCT	AAATCAGACA	CTAACCCCAA	ATGAGAGAAG	TGCCACCATA	3600
ACTGCAGCCT	GAGAGTTTGG	CGATCTCTGG	TATCTCAGTC	AGGTCAATGA	TAGGATGACA	3660
ACAGAGGAAA	GAGAATGATT	CCCCACAGGC	CAGCAGGCAG	TTCCCAGTCT	AGACCCTCAT	3720
TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCTGCAGAC	ATTTGCTAAC	TTGTGTGCTA	3780
GAAGGACTAA	GGAAAAC TAG	GAAGAAGTCT	ATGAATTACT	CAATGATGTC	CACCATAACA	3840
CAGGGAAGGG	AAGAAAATCC	TACTGCCTTT	CTGGAGAGAC	TAAGGGAGGC	ATTGAGGAAG	3900
CGTGCCTCTC	TGTCACCTGA	CTCTTCTGAA	GGCCAACTAA	TCTTAAAGCG	TAAGTTTATC	3960
ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAA	CTTCAAAAGT	CTGCCGTAGG	CCCGGAGCAA	4020
AACTTAGAAA	CCCTATTGAA	CTTGGAACCC	TCGGTTTTTT	ATAATAGAGA	TCAGGAGGAG	4080
CAGGCGGAAC	AGGACAAACG	GGATTAAAAA	AAAGGCCACC	GCTTTAGTCA	TGACCCTCAG	4140
GCAAGTGGAC	TTTGAGGGCT	CTGGAAAAGG	GAAAAGCTGG	GCAAATTGAA	TGCCTAATAG	4200
GGCTTGCTTC	CAGTGCGGTC	TACAAGGACA	CTTTAAAAAA	GATTGTCCAA	GTAGAAGTAA	4260
GCCGCCCCCT	CGTCCATGCC	CCTTATTTCA	AGGGAATCAC	TGGAAGGCCC	ACTGCCCCAG	4320
GGGACAAAGG	TCCTCTGAGT	CAGAAGCCAC	TAACCAGATG	ATCCAGCAGC	AGGACTGAGG	4380
GTGCCTGGGG	CAAGCGCCAT	CCCATGCCAT	CACCCTCACA	GAGCCCTGGG	TATGCTTGAC	4440
CATTGAGGGC	CAGGAGGTTG	TCTCCTGGAC	ACTGGTGCGG	TCTTCTTAGT	CTTACTCTTC	4500
TGTCCCGGAC	AACTGTCCTC	CAGATCTGTC	ACTATCTGAG	GGGGTCCTAA	GACGGGCAGT	4560
CACTAGATAC	TTCTCCCAGC	CACTAAGTTA	TGACTGGGGA	GCTTTATTCT	TTTCACATGC	4620
TTTTCTAATT	ATGCTTGAAA	GCCCCACTAC	CTTGTTAGGG	AGAGACATTC	TAGCAAAAGC	4680
AGGGGCCATT	ATACACCTGA	ACATAGGAGA	AGGAACACCC	GTTTGTGTGC	CCCTGCTTGA	4740
GGAAGGAATT	AATCCTGAAG	TCTGGGCAAC	AGAAGGACAA	TATGGACGAG	CAAAGAATGC	4800
CCGTCCTGTT	CAAGTTAAAC	TAAAGGATTC	CACCTCCTTT	CCCTACCAAA	GGCAGTACCC	4860
CCTCAGACCC	AAGGCCCAAC	AAGGACTCCA	AAAGATTGTT	AAGGACCTAA	AAGCCCAAGG	4920

CCTAGTAAAA	CCATGCAGTA	ACCCCTGCAG	TACTCCAATT	TTAGGAGTAC	AGAAACCCAA	4980
CAGACAGTGG	AGGTTAGTGC	AAGATCTCAG	GATTATCAAT	GAGGCTGTTG	TTCCTCTATA	5040
GCCAGCTGTA	CCTAGCCCTT	ATACTCTGCT	TTCCCAAATA	CCAGAGGAAG	CAGAGTGGTT	5100
TACAGTCCTG	GACCTTCAGG	ATGCCTTCTT	CTGCATCCCT	GTACATCCTG	ACTCTCAATT	5160
CTTGTTTGCC	TTTGAAGATA	CTTCAAACCC	AACATCTCAA	CTCACCTGGA	CTATTTTACC	5220
CCAAGGGTTC	AGGGATAGTC	CCCATCTATT	TGGCCAGGCA	TTAGCCCAAG	ACTTGAGCCA	5280
ATCCTCATAC	CTGGACACTT	GTCCTTCGGT	AGGTGGATGA	TTTACTTTTG	GCCGCCCATT	5340
CAGAAACCTT	GTGCCATCAA	GCCACCCAAG	CGCTCTTCAA	TTTCCTCGCT	ACCTGTGGCT	5400
ACATGGTTTC	CAAACCAAAG	GCTCAACTCT	GCTCACAGCA	GGTTACTTAG	GGCTAAAATT	5460
ATCCAAAGGC	ACCAGGGCCC	TCAGTGAGGA	ACACATCCAG	CCTATACTGG	CTTATCCTCA	5520
TCCCAAAACC	CTAAAGCAAC	TAAGGGGATT	CCTTGGCGTA	ATAGGTTTCT	GCCGAAAATG	5580
GATTCCCAGG	TATGGCGAAA	TAGCCAGGTC	ATTAAATACA	CTAATTAAGG	AAACTCAGAA	5640
AGCCAATACC	CATTTAGTAA	GATGGACAAC	TGAAGTAGAA	GTGGCTTTCC	AGGCCCTAAC	5700
CCAAGCCCCA	GTGTTAAGTT	TGCCAACAGG	GCAAGACTTT	TCTTCATATG	TCACAGAAAA	5760
AACAGGAATA	GCTCTAGGAG	TCCTTACACA	GATCCGAGGG	ATGAGCTTGC	AACCTGTGGC	5820
ATACCTGACT	AAGGAAATTG	ATGTAGTGGC	AAAGGGTTGA	CCTCATTGTT	TACGGGTAGT	5880
GGTGGCAGTA	GCAGTCTTAG	TATCTGAAGC	AGTTAAAATA	ATACAGGGAA	GAGATCTTAC	5940
TGTGTGGACA	TCTCATGATG	TGAATGGCAT	ACTCACTGCT	AAAGGAGACT	TGTGGCTGTC	6000
AGACAACTGT	TACTTTAAAT	GTCAGGCTCT	ATTACTTGAA	GGGCCAGTGC	TGCGACTGTG	6060
CACTTGTGCA	ACTCTTAACC	CAGCCACATT	TCTTCCAGAC	AATGAAGAAA	AGATAAAACA	6120
TAACTGTCAA	CAAGTAATTT	CTCAAACCTA	TGCCACTCGA	GGGGACCTTT	TAGAGGTTCC	6180
TTTGACTGAT	CCCGACCTCA	ACTTGTATAC	TGATGGAAGT	TCCTTTGTAG	AAAAAGGACT	6240
TCGAAAAGTG	GGGTATGCAG	TGGTCAGTGA	TAATGGAATA	CTTGAAAGTA	ATCCCCTCAC	6300
TCCAGGAACT	AGTGCTCAGC	TAGCAGAACT	AATAGCCCTC	ACTTGGGCAC	TAGAATTAGG	6360
AGAAGAAAAA	AGGGCAAATA	TATATACAGA	CTCTAAATAT	GCTTACCTAG	TCCTCCATGC	6420
CCATGCAGCA	ATATGGAAAG	AAAGGGAATT	CCTAACTTCT	GAGAGAACAC	CTATCAAACA	6480
TCAGGAAGCC	ATTAGGAAAT	TATTATTGGC	TGTACAGAAA	CCTAAAGAGG	TGGCAGTCTT	6540
ACACTGCCGG	GGTCATCAGA	AAGGAAAGGA	AAGGGAAATA	GAAGAGAACT	GCCAAGCAGA	6600
TATTGAAGCC	AAAAGAGCTG	CAAGGCAGGA	CCCTCCATTA	GAAATGCTTA	TAAAACAACC	6660

CCTAGTATAG	GGTAATCCCC	TCCGGGAAAC	CAAGCCCCAG	TACTCAGCAG	GAGAAACAGA	6720
ATGGGGAACC	TCACGAGGAC	AGTTTTCTCC	CCTCGGGACG	GCTAGCCACT	GAAGAAGGGA	6780
AAATACTTTT	GCCTGCAACT	ATCCAATGGA	AATTACTTAA	AACCCTTCAT	CAAACCTTTC	6840
ACTTAGGCAT	CGATAGCACC	CATCAGATGG	CCAAATCATT	ATTTACTGGA	CCAGGCCTTT	6900
TCAAAACTAT	CAAGCAGATA	GTCAGGGCCT	GTGAAGTGTG	CCAGAGAAAT	AATCCCCTGC	6960
CTTATCGCCA	AGCTCCTTCA	GGAGAACAAA	GAACAGGCCA	TTACCCTGGA	GAAGACTGGC	7020
AACTGATTTT	ACCCACAAGC	CCAAACCTCA	GGGATTTTCA	TATCTACTAG	TCTGGGTAGA	7080
TACTTTCACG	GGTTGGGCAG	AGGCCTTCCC	CTGTAGGACA	GAAAAGGCCC	AAGAGGTAAT	7140
AAAGGCACTA	GTTTCATGAA	TAATTCCCAG	ATTCGGACTT	CCCCGAGGCT	TACAGAGTGA	7200
CAATAGCCCT	GCTTTCAGG	CCACAGTAAC	CCAGGGAGTA	TCCCAGGCGT	TAGGTATACG	7260
ATATCACTTA	CACTGCGCCT	GAAGGCCACA	GTCCTCAGGG	AAGGTCGAGA	AAATGAATGA	7320
AACACTCAAA	GGACATCTAA	AAAAGCAAAC	CCAGGAAACC	CACCTCACAT	GGCCTGCTCT	7380
GTTGCCTATA	GCCTTAAAAA	GAATCTGCAA	CTTTCCCCAA	AAAGCAGGAC	TTAGCCCATA	7440
CGAAATGCTG	TATGGAAGGC	CCTTCATAAC	CAATGACCTT	GTGCTTGACC	CAAGACAGCC	7500
AACTTAGTTG	CAGACATCAC	CTCCTTAGCC	AAATATCAAC	AAGTTCTTAA	AACATTACAA	7560
GGAACCTATC	CCTGAGAAGA	GGGAAAAGAA	CTATTCCACC	CTTGTGACAT	GGTATTAGTC	7620
AAGTCCCTTC	CCTCTAATTC	CCCATCCCTA	GATACATCCT	GGGAAGGACC	CTACCCAGTC	7680
ATTTTATCTA	CCCCAACTGC	GGTTAAAGTG	GCTGGAGTGG	AGTCTTGGAT	ACATCACACT	7740
TGAGTCAAAT	CCTGGATACT	GCCAAAGGAA	CCTGAAAATC	CAGGAGACAA	CGCTAGCTAT	7800
TCCTGTGAAC	CTCTAGAGGA	TTTGCGCCTG	CTCTTCAAAC	AACAACCAGG	AGGAAAGTAA	7860
CTAAAATCAT	AAATCCCCAT	GGCCCTCCCT	TATCATATTT	TTCTCTTTAC	TGTTCTTTTA	7920
CCCTCTTTCA	CTCTCACTGC	ACCCCTCCA	TGCCGCTGTA	TGACCAGTAG	CTCCCTTAC	7980
CAAGAGTTTC	TATGGAGAAT	GCAGCGTCCC	GGAAATATTG	ATGCCCCATC	GTATAGGAGT	8040
CTTTCTAAGG	GAACCCCCAC	CTTCACTGCC	CACACCCATA	TGCCCCGCAA	CTGCTATCAC	8100
TCTGCCACTC	TTTGCATGCA	TGCAAATACT	CATTATTGGA	CAGGAAAAAT	GATTAATCCT	8160
AGTTGTCCTG	GAGGACTTGG	AGTCACTGTC	TGTTGGACTT	ACTTCACCCA	AACTGGTATG	8220
TCTGATGGGG	GTGGAGTTCA	AGATCAGGCA	AGAGAAAAAC	ATGTAAAAGA	AGTAATCTCC	8280
CAACTCACCC	GGGTACATGG	CACCTCTAGC	CCCTACAAAG	GACTAGATCT	CTCAAAACTA	8340
CATGAAACCC	TCCGTACCCA	TACTCGCCTG	GTAAGCCTAT	TTAATACCAC	CCTCACTGGG	8400
CTCCATGAGG	TCTCGGCCCA	AAACCCTACT	AACTGTTGGA	TATGCCTCCC	CCTGAACTTC	8460
AGGCCATATG	TTTCAATCCC	TGTACCTGAA	CAATGGAACA	ACTTCAGCAC	AGAAATAAAC	8520

ACCACTTCCG	TTTtagtagg	ACCTCTTGTT	TCCAATCTGG	AAATAACCCA	TACCTCAAAC	8580
CTCACCTGTG	TAAAATTTAG	CAATACTACA	TACACAACCA	ACTCCCAATG	CATCAGGTGG	8640
GTAACtCCTC	CCACACAAAT	AGTCTGCCTA	CCCTCAGGAA	TATTTTTTGT	CTGTGGTACC	8700
TCAGCCTATC	GTTGTTTGAA	TGGCTCTTCA	GAATCTATGT	GCTTCCTCTC	ATTCTTAGTG	8760
CCCCCTATGA	CCATCTACAC	TGAACAAGAT	TTATACAGTT	ATGTCATATC	TAAGCCCCGC	8820
AACAAAAGAG	TACCCATTCT	TCCTTTTGTT	ATAGGAGCAG	GAGTGCTAGG	TGCACTAGGT	8880
ACTGGCATTG	GCGGTATCAC	AACCTCTACT	CAGTTCTACT	ACAAACTATC	TCAAGAACTA	8940
AATGGGGACA	TGGAACGGGT	CGCCGACTCC	CTGGTCACCT	TGCAAGATCA	ACTTAACTCC	9000
CTAGCAGCAG	TAGTCCTTCA	AAATCGAAGA	GCTTTAGACT	TGCTAACCGC	TGAAAGAGGG	9060
GGAACCTGTT	TATTTTTTAGG	GGAAGAATGC	TGTTATTATG	TTAATCAATC	CGGAATCGTC	9120
ACTGAGAAAAG	TTAAAGAAAT	TCGAGATCGA	ATACAACGTA	GAGCAGAGGA	GCTTCGAAAC	9180
ACTGGACCCT	GGGGCCTCCT	CAGCCAATGG	ATGCCCTGGA	TTCTCCCCTT	CTTAGGACCT	9240
CTAGCAGCTA	TAATATTGCT	ACTCCTCTTT	GGACCCTGTA	TCTTTAACCT	CCTTGTTAAC	9300
TTTGTCTCTT	CCAGAATCGA	AGCTGTAAAA	CTACAAATGG	AGCCCAAGAT	GCAGTCCAAG	9360
ACTAAGATCT	ACCGCAGACC	CCTGGACCGG	CCTGCTAGCC	CACGATCTGA	TGTTAATGAC	9420
ATCAAAGGCA	CCCCTCCTGA	GGAAATCTCA	GCTGCACAAC	CTCTACTACG	CCCCAATTCA	9480
GCAGGAAGCA	GTTAGAGCGG	TCTCGGCCAA	CCTCCCCAAC	AGCACTTAGG	TTTTCCTGTT	9540
GAGATGGGGG	ACTGAGAGAC	AGGACTAGCT	GGATTTCCTA	GGCTGACTAA	GAATCCCTAA	9600
GCCTAGCTGG	GAAGGTGACC	ACATCCACCT	TTAAACACGG	GGCTTGCAAC	TTAGCTCACA	9660
CCTGACCAAT	CAGAGAGCTC	ACTAAAATGC	TAATTAGGCA	AAGACAGGAG	GTAAAGAAAT	9720
AGCCAATCAT	CTATTGCCTG	AGAGCACAGC	AGGAGGGACA	ATGATCGGGA	TATAAACCCA	9780
AGTCTTCGAG	CCGGCAACGG	CAACCCCTT	TGGGTCCCCT	CCCTTTGTAT	GGGAGCTCTG	9840
TTTTCATGCT	ATTTCACTCT	ATTAAATCTT	GCAACTGCAC	TCTTCTGGTC	CATGTTTCTT	9900
ACGGCTTGAG	CTGAGCTTTC	GCTCGCCATC	CACCACTGCT	GTTTGCCGCC	ACCGCAGACC	9960
CGCCGCTGAC	TCCCATCCCT	CTGGATCATG	CAGGGTGTCC	GCTGTGCTCC	TGATCCAGCG	10020
AGGCACCCAT	TGCCGCTCCC	AATCGGGCTA	AAGGCTTGCC	ATTGTTCCCTG	CATGGCTAAG	10080
TGCCTGGGTT	CATCCTAATT	GAGCTGAACA	CTAGTCACTG	GGTTCCATGG	TTCTCTTCTG	10140
TGACCCACAG	CTTCTAATAG	AGCTATAACA	CTCACC GCAT	GGCCCAAGGT	TCCATTCCCTT	10200
GAATCCATAA	GGCCAAGAAC	CCCAGGTCAG	AGAACACGAG	GCTTGCCACC	ATCTTGGGAG	10260
CTCTGTGAGC	AAGGACCCCC	AAGTAACACA	ACCATGAGGG	TGCAAATGCA	TGGGCCACTA	10320

ATGGTAGAGC AAGAAAACAG AAGGGCCCTG GTTCCTCGAA GGCATCAGTG AGCTGAAATG 10380
CCTGCCCTGG ATGTCCTATT CCTAGGTGTT TTTCTGCCTG AAGCAGATTA AACCTTTGT 10440
TCACTTCTCC AAGTAGGGCT TCTATTACAG CCCAAATCAA TCCCCACCCC AGATGACAT 10499

(2) INFORMATION FOR SEQ ID NO: 4: HE2

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2784 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE : DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CTCCTTCAGGAGAACAAAGAACAGGCCACTACCCAAGAGAAGACTGGCAACTAGATTTTACCCATATGCCCAAATCTCAG
GGATTTTCAGTATCTACTAGTTTGGGTAGATACTTTCACTGGTTGGGCAGAGGCCTTCCCCTGTAGGACAGAAAAGGCCCA
AGAGGTAATAAACGTTTCATGAAATAATTCCCAGATTTCGGACTTCCCCAAGGCTTACAGAGTGACAATGGCCCTGCTTTCA
AGGCTACAGTAACCCAAGGAGTATCCCAGGTGTTAGGTATACAATATCACTCACACTGCGCCTGGAGGCCACAGTCCTCA
GGAAAGGTGGAGAAAATGAACAAAACACTCAAATGACATCTAAAAAAGCTAATCCAGGAAACCCACCTCGCATGGCCTGC
TCTGTTGCCTATAGCCTTACTAAGAATCCGAAACTCTCCCCAAAAAGCAGGACTTAGTCCATACAAAATGCTGTATGGAC
GGCCCTTCTTAACCAATGAACCTTGGGCTTGACCGAGAGACAGCCAACTTAGTTGCAGACATCATCTCCTTAGCCAAATAT
CAACAGGTTCTTAAACATTACAGGGAGCCTGTCCCCAAGAAGAGGGAAAGGAACTATTCCACCCTGGTGACATGGTATT
AGTCAAGTCCCTTCCCCTCTAATTCCCCATCCCTAGATACATCCTGGGAAGGAAACTACCCAGCCATTTTATCTACCCTAA
CGGCAGTTAAAGTGGCTGGAGCGGAGTCTTGGATACATCACACTCAAGTCAAACCCTGGATACTGCCAAAGGAACTCAAA
AATCCATGAGACAATGCTAGCTATTCTGTGAACCTCTAGAGGATCTGCGCCTGCTCTTCAAATGACAACAGGGGGAAA
GTAATAAAATCGTAAATCCCCTGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCTTTCACTCTCAC
TGCACCCCGTCCATGCCACTGCACCCCGTCCATGCCCCGTCCATGCCAGTAGCTCCCCTTAGCAAGAGTTTCTATGGAGA
ATGCAGCGTCCCGGAAATATTGATGCCCCATTGTATAGGAGTTTATCTAAGGGAACCCCCACCTTCACTGCCACACCCA
TATGCCCCACAACCTGCTATAACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAACGATTAATC
CCAGTTGTCTTGGAGGACTTGGAGGACTCACTTCACTCATACCAGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCCAA
CAGAAAAACACATAAAGGAAGTAATCTCCCAACTGACCTGGGTACATAGCACCCCTGGCCCCCTACAAAGGACTAGATCTC
TCAAAACATACATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTATTTAATACCACCCTGACTGGGCTCCATGAGGT
CTCGGCCCCAAAACCCCTACTAAGTGTGGATGTGCCTCCCCCTGCACCTTAGGCCATACATTTCAATCCCTATACCTGAAC
AATGGAACAACCTTCAGCACAGAAATAAACACCCTTCTGTTTTAGTAGGTCTCTTTCCAATCTGGAAATAACCCATACC
TCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACCTCCAATGCATCAGGTGGGTAACTCCTCCCAC
ACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAAT
CTGTGTGCTTCTCTCATTCTTAGTGGCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAAG
CCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTATTGGAGCAGGAGTGCTAGGCGGAGTAGCTACTGGCATTGGCGG
TATCACAACCTCTACTCAGTTCTACTACAACTGTCTCAAGAACTAAATGGTGACATGGAATGGGTGCTGATACCCCTGG
TCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAA
AGCGGGGGAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAA
AGAAATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGC
CCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTT
GTTAAGTTTGTCTTTTCCAGAATCGAAGCAGTAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAG
TAAATCTACCACGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCAGG
AAATCTCAACTGCACAACCTCTACTACGCCCAATTGAGCAGGAAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACA
GCAGTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGACAGGAATAACTAGATTTCTAGACCAACTAAGAATCCCTAAG
ACTAGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCAACCAATCAGATACTAA
GAGAGCTCACTAAATGCTAATTAGGCAAAAACAGGAGATAAAGAAATAGCCAATCATCTGTTG

(2) INFORMATION FOR SEQ ID NO: 5: HE3

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1799 base pair

(B) TYPE: nucleotide
(C) STRANDS NUMBER: single
(D) CONFIGURATION: linear

(ii) MOLECULE TYPE : DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGGATTCCTTAGTCGGCCTAGGAAATCCAGCTAATCCTGTCTCTCAGTCCCCCACTCAACAGGAAAACCCAAGTGCTGTT
GGGGAGGTTGGCTGACGACCAGTCTAACTGCTTCCTGCGGAATTGGGGCATAGTAGGGGTTGTGCAGTTGAGATTTCCCTC
GGGAGGGGTGCGTTCGATATCATTACAATTGGAGCATGGGCTAGTAGGCCGGTCCAGGGGTCCACGGTAGATCCTTAGTCA
TGGACTTCATCTGGGGTTCCATTTGAAGAACGATTTGTAGCTTTACAACCTTTGATTCTGGAAGAGACAACTTAACAAGG
AGGTTAAAGATACAGGGTCCAAAGAGGAGTATCAATATTAGAGCTGCTAGAGATCCTAAGAAGGGGAGAATCCAGGGCAT
CCATTGGCTGAGGAGGCCCCAGGGTCTGGTGTCTTTGAAGCTCCTCTGTTCTACGTTGTATTCAATCTCGAATTTCTTCA
ACTTTCTCTGTGACAATTCAGGATTGATTAACATAATAACAACATTCTTCCGCTAAAATAACATAATAACAACATTCTTC
CCCTAAAAATAAACAGCTTCCCCCTCTTTTCAAGAGGTTAGCAAGTCTAAAGCTCTTCAATTTTGAAGGACTACTGATGCTA
GGAAGTTAAGTTGATCTTGCAAGGTGACCAGGGAGTCGGCAACCCATTCCATGTCACCATTGAGTTCTTGAGATAGTTTG
TAGTAGAACTGAGTAGAGTTGTGGTACCGCCAATGCCAGAACCTAGTCCACCTAGCACTCCTGCTCCGATAACAAAAGG
AAGAATGAGTACTCTTTTGTGTGGGGCTTAGGTACAACATAATGTATAAATCTTGTTCAGTGTAATGGTCATGGGGG
CACTAAGAATGAGAGGAAGCACATAGATTCTGAAGAGCCATTCAAACAACGATAGGCTAAGGTACCACAGACAAAAATA
TTCCTGAGGGTAGGCAGACTATTCGTGTGGGAGGAGTTACCCACCTGATGCATTGGGAGTTGGTTGTGTCTACAGTATTG
CTAAATTTTACACAGGTGAGGTTTGGGTATGGGTATTTCCAGATTGGAAACAAGAGGTCCTACTAAAACGGAAGTGGT
GTTTATTTCTGTGCTGTAGTTGTTCCATTGTTTCAAGTACAGGGATTGAAATGCATGGCCTGAAATACAGGGGGAGGCACA
ACCAACAGTTAGTAGGGTTTTGGACCGAGACCTCATGGAGCCAGTGAGGGTGGTATTAAATAGGCTTACCAGGCAAGTA
TGGGTATGGAGGGTTTCATGTAGTTTTTAAGAGATCTAGTCCTTTGTAGGGGGCTAGGGGTGCTATGTACCCGGGTCAGTTG
GGAGGTTACTTCCTTTACATGTTTTTCTCTTGCCTGATCTTGAACCTCCACCCCCCTCAGACATACCAGTATGGGTGAAGT
AAGTCCGACAGACAGTGGCTCCAAGTCTTCCAGGACAACCTAGGATTAATCATTTTCCCTGTCCAATAATGAGTATTTGCA
TGCATGCAAAGAGTGGCAGAGTTATAGCAGTTGTGGGGCATATGGGTGTGGGCAGTGAAGGTGGAGTTTCCTTTAGGTAA
ACTCCTATTTGATGGGGCATCAATATTTCTGGGAAGCCGCATTCTTCATAGAACTCTTGGTAAGGGGAGCTGCTGGTTG
TACAGCAGCATGGAGGGGGTGCAGTGAGAGTGAAAGGGGGTAAGAGAACAGTAAAGAGAAAAATATGATAAGGGGAGGGCC
ATGGGGATTACGATTTTAGTTACTTTTCTCCTCACGGTTGT

(2) INFORMATIONS FOR SEQ ID NO: 6: HG3

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1489 base pair
(B) TYPE: nucleotide
STRANDS NUMBER: single
(D) CONFIGURATION: linear

(ii) MOLECULE TYPE : DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TGGTGCTTGC CCCGGGCACT CTCAGTCCTG CTGCTGGATC ATCTGGTTAG TGGCTTCTGA	60
CTCAGAGGAC CTACGTCCCC TGGGGCAGTG GGCCTTACAG TGATTCCTT GACACGAGGT	120
GCATGGACGA GGGGGCGGCT TATTTCTATT TGGACAATCT TTTTAAAGT GTCCTTGTAG	180
ACCGCACTGG AAGCAAACCC TATTAGGCAT TTGATTTGCC TAGCTTTTCC CTTTCCAGT	240
GCCTCCAAAG TCCGCTTGCC TGAGGGCCAT GACTAAAGCG GTGGCCTTTT TTTTATCCCA	300
TTTGTCCCAT TCTGCCTGCT CATCCTGATC TCTATTATAA AAAACTGAGG TTGCCAAGTT	360
CAATAGGGTT TCTAAGTTTT GTTCCGGGCC TAAGGCAGAC TTTTGAAGTT TTTTCCTAAT	420

GTCTGTAGCT GACTGAGTGA TAAACTTATC CTTTAAGATT AGTTGGCCTT CAGTAGAGTC	480
AGTTGACAGA GAGAGGTATG CTTCTCAAT GCCTCCGTTA GTCCTCCAG AAAGGCGGTA	540
GGATTTTCTT CCTTCCCTG TGTATAGTG GACATCATTG AATAACTCAC AGGCTTCTTT	600
CTAGTTTTC TTAGTCCTTC TAGCACGCAA GTTAGCAAAT GTCTGCGGCA CCAATCTCCA	660
TGTTCTGATT CTGTGTCCCA GTGAGGGTCT AACTGGGAA CTGCCTGCTG GCCTGTGGGG	720
AATCGTTCTC TTTCTCTGT TGTCGACCTA TCATTGACCT GACTGAGATA CCAGAGATCG	780
CCAAACTCTC AGGCTGCAGT TACGGCGACA CTTCTGTCAT TTGGGGTTAG TGTCTGATTT	840
AGCAGTAACA TTATATCTCT CCATATCAGA TCAAAGGATT GTCCTAAACC TTGTAAAACA	900
TCAATATAGC CATTAGGGTT ATCTGAGAAT TTACCTAGGT CTATTTTAAT TTAAAGTCTG	960
GGAGAGAAAA AGGCACATGC ACTCTGGCTG GGCCGAATTC TCTTCCTCCC ACTGCGTCTG	1020
AGAGAGAAAA AGGTACGTGC ACTCTGGCTG GGCCGAATTC TCCTCCCACC GCTTGGAGGG	1080
GGCATAATCG GGGAATATTG GCATTCTTTG GTTAGTTGTT TACCCCTTTG TCTATCTCCT	1140
TTTGGACCGT TTGGGTTGAA GGGGGGTCCT TATTATTTGG GGAAGGAGTC TGGGGGATGC	1200
TGGGGTAGGG AGGTAGACTC TGAGGGCTTC CTGTAGGGCA TAAATCACAC TTTTACATA	1260
ATTGCGAGTT GTCTCTTAAT GAAAAGAAAG TTTGTACGTA TGACACTTCA CACCATTTC	1320
CTTCTTTTCT ACAAAGAGG TCTAGCTGTA AGATGGTGTT ATAATTTATG CTTCCCTCAG	1380
GATGCCAGGT TTCTCCCCCT TAAAGAGTAT ATCGTTGCCA GGCGGTACTG CAGAAGAATA	1440
TGTCTTTTTT TTCTTAGCAT CTGAGAGTCA AATTGGTCCC AATTCTCCA	1489

(2) INFORMATION FOR SEQ ID NO: 7: HE4

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1216 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 7:

TAAAGATACA GGGATTGAAA TGTATGGCCT GAAGTGCAGG GTCATATAGG TGTGGGTGGT	60
GAAAATGGGG TTTCTTTAG AAAAATCCT ATACGATGGG TCATCAATAT TTCCAGGAAG	120
CCGCATTCTC CATAGAAGCT CTTGGTAATG GGAGCTACTG GTAGTACAGT GGCATGGAGG	180
GGGTGCAGTG AGAGTGAAAG AGGGTAAAAG AACAGTAAAG AGAAAAATAT GATAAGGGAG	240
GGGTTCACTG AGAGTGAAAG GGGGTAAAG AACAGTAAAG AAAAAATAT GACAAGGAGG	300
GCCATGAGGA TCTACGATTC TAGTTACTTT CCTCACGGTT GTCGCTTGAA GAGCAGGTGC	360

AGATCCTCTA GAGGTTTACA GGAATAGCTA GCGTTGTCTC CTGGATTTTC GGGTTCCTTT	420
GGCAGTATAC AGAGTTTGAC TCGAGTGTGA TGTATTCAAG ACTCCACTCC AGCCACTTTA	480
ACCGCAGTTG GGGTAGATAA AATGACTGGG TAGGGTCCTT CCCAGGATGT ATCTAAGGAT	540
GGGGACTTAG AAGGAAGGGA CTTGACTAAT ACCATGTCAC CAGGGTGCAA TAATTACTTT	600
CCCTCTTCTC GGGAACAGGT TCCCTGTAAT GTTTTAAGAA CTTGTTGATA TTTGGCCAAG	660
GAGGTGATGT CTGCAACTAA GCTGGCCATC TCTCGGTCAA GCACAAGGTC CTTGGTTAGG	720
AAGGGCCATC CATAACAGCAT TTTGTATGGG CTAAGTCCTG CTTTTTGGGG AGAGTTTGG	780
ATTCTTAGTA AGGCTGTAGG CAACAGAGCA GGCCATGCAA GGTGGGTTTC TTGGGTTAGC	840
TTTTTTAAAT GTCGTTTGAG TGCTTCATTC ATTTTCTTGA CTTTTCCTGA GGATTGTGGC	900
CTCCACGCGC AGTGTAAGTG ATATTGTATG CCTAATGCCT GGGATACTCC CTGGGTTACT	960
GTAGCCTTGA AAACGGGGCC ATTGTCACTC TGTAAGCCTC GGGGAAGTCC GAATCTGGGA	1020
ATTATTTTCAT GAATTAGTGC CTTTATTACA TCTTGGTCCT TTTCTGTCCT ACAAAGGAAG	1080
GCCTCTGCCC AACCAGTGAA AATATCTACC CAGACTAGTA GATACTGAAA TCCCTGAGAT	1140
TTGGGCATGT GGGTAAAATC TAGTTGCCAG TCTTCTCCTG AGTAATGGCC TGTTCCTTGT	1200
TCTCCTGAAG GAGCTT	1216

(2) INFORMATION FOR SEQ ID NO: 8: HE5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 976 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

AGTGATAATG GAATACTTGA AAGTAATCCC CTCCTCCAG GAACTAGTGC TGAGCTGGCC	60
AAACTAATAG CCCTCACTCG GGCCTAGAA TTAGGAGAAG AGAAAAGGGT AAATATATAT	120
ACAGACTATA AGTATGCTTA CCTAGTCCTT CATGCCCATG CAGCAATATG GAGAGAAAGG	180
GAATTCCTAA CTTCCAAAGG AACACCTATC AAACATCAGG AAGCCATTAG GATATTATTA	240
TTGGTGGTAC AGAAACCTAA AGAGGTGGCA GTCCTACACT GCTGGGGTCA TCAGAAAAAA	300
AAGGAAAGGG AAATAGAAGG GAACTACCAA GCAGATATTG AAGCCAAAAG AGCCGCAAGG	360
CAGGACCCTC CATTAGAAAT GCTTATAGAA GGACCCCTAG TGTGGGGTAA CCCCCTCCAG	420
GAAAGCAATC CCCAGTACTC AGCAGGAGAA ATAAAATGGA GAACCTCACG AGGACATACT	480
TTCCTCCCCT CAGGATGGCT AGCCACCAAA GAAGGAAAAA TGCTTTTGCC TGCAGCTAAC	540

CAATGGAAAT TACTTAAAC CCTTCACCAA ACCTTTCACT TAGGATTGAT AGCACCCATC	600
AGATGGCCAA ATTATTATTT ACTGGATCAG GCCTTTTCAA AACTATCAAG CAGGTAGTCA	660
GGGCTGTAA AGTGTGCCAA AGAAATAATC TCCTGCACTG CAAGCCATAC ATTTCAATCC	720
CTGTATCTTT AACCTCCTTG TTAAGTTTGT CTCTTCCAGA ATCAAAGCTG TAAAACTACA	780
AATGGTTCTT CAAATGGAGT CTCAGATGCA GTCCATGACT AAGATATACC GCAGCCCCCT	840
GGAGGGGGCC TGCTAGCCCA TGCTCCAATG TTAATGACAT CGAAGGCACC CCTCCCGGGG	900
AAATCTCAAC TGCACAACCC CTACTATGTC CCAATTCAGC AGGAAGCAGT TAAAGCGGTC	960
ATCGGCCAAC CTCCCC	976

(2) INFORMATIONS FOR SEQ ID NO: 9: HE6

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 942 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AGAGGAGAAC AGCAGCATAA GCGGCTGGCA GAGGTAGGGA AAGACCAGCA AGAAGAAAAG	60
AGAGAAAGAG AAAGAGAAAG TCAGAGAAAG AGACAGAGAG AGGAAGAGAC AAAGAGACAG	120
AAAGTCAAAG AGGTAGTAGT CAGAAACAGA GACAAAAAAA AGGAGTCAGA AAGAGGGACA	180
GACACAGAAA GTCAAAAAAA AAGTTAAGAA GAAAGGAAAA GACAAAGAAG AAGTCGAAGA	240
GGAGAAAGAG AGAGATAGAA GTAGTAAAGA AAAAAACAGC ATATCCCATT CCTTTAAAGC	300
CAGGGTAAAT TTCTATCTAC CCAGCCAAGG CATATTCTAC TTATGTGGAT CTTCAACCCA	360
TATCTGCCTC TCAGACAGTT TGCAAGAAAT AATGAAATCT ATCCTTACTT TACAATCCCA	420
AATAGACTCT TTGGCAGCAG TGA CTCTCCA AA ACTGCAGA GGCCTAGACC TCCTCACTGC	480
TGAAAAAGGA GGACACTACA CCTTCTTAGG GGAAGAATGT TGTTTTTACA CTAACCAGTC	540
GGGGATAGTA TGAGATGCTG CCCGGAGTTT ACAGGAAAAG GCTTCTGAAA TCAGACAACG	600
CCTTTCAAAT TCTTATACCA ACTTCTGGAG TTAGGCAACA TGGCTTCTCC CCTTTCTAGG	660
TCCTGTGGCA GCCATCTTGC TGTTACTCGC CTTTGGGCCC TGTATTTTTA ACCTTCTTGT	720
CAAATTTGTT TCCTCTAGAA TCGAGGCCAT CAAGCTACAG ATGGTCTTAC AAATGGAACC	780
CCAAAAGAGT TCAACTAACA ACTTCTACCG AGGACCCCTG GATCAACCCA CTGGCACTTC	840
CCCTGGCCTA GAGAGTTCCC CTCTGAAGGA CACCGCAACT GCAGGGCCCT TCTTTGCCCC	900
ATCCAGCAGG AGTAGCTAGA GTGGTCATCG GCCAAATTGC CA	942

(2) INFORMATION FOR SEQ ID NO: 10: HG6

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1375 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCCCAATATT CTCTTTCTGA TGGGGAAAAA TGGCCACCTG AGGGAAGCAC AAATTACAAT	60
ACTATCCTGC AGCTTGATCT TTTCTGTAAG AGGGAAGGCA AATGGAGTGA AATACCTTAT	120
GTCCAAGCTT TCTTTTCATT GAGGGAGAAT ACACAACTAT GCAAAGCTTG CAATTTACAT	180
CCCACAGGAG GACCCCTCAG CTTACCCCCA TATCCTAGCC TCCCTATAGC TTCCCTTCCT	240
ATTGATGATA CTCCTCCTCT AATCTCCCCT GCCCAGAAGG AAATAAGCAA AGAAATCTCC	300
AAAGGTCCAC AAAAACCCCC GGGCTATCGG TTATGTCCCC TTCAAGCTGT AGGGGGAGGG	360
GAATTTGGCC CAACCCGGGT GCATGTCCCC TTCTCCCTCT CTGATTTAAA GCAGATCAGG	420
CAGACCTGGG GAAGTTTTCA GATGATCCTG ATAGGTACAT AGATGTCCTA CAGGGTCTAG	480
GGCAAACCTT TGACCTCACT TGGAGAGACG TCATGCTACT GTTAGATCAA ACCCTGGCCT	540
TTAATGAAAA GAATGCGGCT TTAGCTGCAG CCTGAGAGTT TGGAGATACC TGGTATCCTA	600
GTCAAGTAAA TGAAAGAATG ACAGCCGAAG AAAGGGACAA CTTCCCTACT GGTCAGCAAG	660
CCATCCCCAG TATGGATCCC CACTGGGACT TTGACTCAGA TCATGGGGAC TGGAGTCGTA	720
AACATCTGTT GATCTGTGTT CTGGAAGGAC TAAGGAGAAT TGGGAAAAAG CCCATGAATT	780
ATTCAATGAT ATCCACCATA ACCCAGGGAA AGGAAGAAAA TCCTTCTGCC TTCCTCGAGC	840
GGCTACAAGA GGCCTTAAGA AAATATACTC CCCTGTCACC CGAATCACTC GAGGGTCAAT	900
TGATTCTAAA AGATAAGTTT ATTACCCAAT CAGCCACAGA TATCAGGAGA AAGCTCCAAA	960
AGCAAGCCCT GAGCCCTGAA CAAAATCTAG AGACATTATT AAACCTGGCA ACCTTGGTGT	1020
TCTATAATAG GGACCAAGAG GAACAGGCCC AAAAGGAAAA GCGAGATCAG AGAAAGGCCG	1080
CAGCCTTAGT CATGGCCCTC AGACAAACAA ACCTTGGTGG TTCAGAGAGG TCAGAAAATG	1140
GAGCAGGCCA ATCACCTGGT ACGGCTTGTT ATCAGTGCGG TTTACTAGGA CACTTTAAAA	1200
AAGATTGTCC AATAAGAAAC AAGCTGCCCC CTCATCCGTG TCCACTATGC CGAGGCAATC	1260
ACTGGAAGGT GCACTGCCCC AGAGGATGAA GGTTCCCTGG GTTAGAAGCC CCCAACCAGA	1320
TGATCCAACA ACAGGACTGA GGGTGCCCGG GGCAAGCACC AGCTCATGTC ATCAC	1375

(2) INFORMATION FOR SEQ ID NO: 11: HE7

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 944 base pair
 (B) TYPE: nucleotide
 (C) STRANDS NUMBER: single
 (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ACCTAGGAGG AACTGTCTTC AGGACAGGAC TATAGATGCT TCCTCCCAGG CGATTAAGGG	60
AAAAAGACAC AATGGGTATT CAGTAAGTGA TAAGGAAACT CTTGTAGAAG CAGAGTTAGG	120
AAAATTGCCT AATAATTGGT CTGCTCAAAT GTGCGAGCTG TTTGCACTCA GCCAAACCTT	180
AAAAGTATTA CAGAATCAGG AAGAAGCCAT CTATACCAAT TCTAAGTTAA TATGGACTGA	240
ACGAGAACTT ATTAATAGCA AAGAATAATT GAAATCCCAA ACTTACAAGG TTTTCAACAA	300
AAGCACAGTT TGCTAAAAGT TAACTGTGTA ACATGTATTA TCCTACTACC ACAAACTCTC	360
AAATGATTTT TCAGACAGTT TGCAAGAAAC AATGAAACCT ATCCTTACTC TACAATCCCA	420
AATAGACTCT TTGGCAGCAG TGA CTCTCCA AAACCACCAA GGCCTAGACC TCCTCACTGC	480
TGAGAAAGGA GGACTCTGCA CCTTCTTAGG GGAAGATTGT TGTTTTTACA CTAACCAGTC	540
AGGGATAGTG TGAGATGCCA CCCAGCGTTT ACAGGAAAAG GCTTCTGAAA TCAGACACAA	600
TGCTTTTCAA ACCTTATAGC AACCTCTGGA GTTCGGCGAC TGGCTTTTCC CCTTTCTAGG	660
TCCTGTGACA GCCATCTTGC TATTACTCGC CTTCGGGCCC TGTATTTTAA ACCTCCTCGT	720
CAAATTTGTT TCCTCTAGGA TCGAGGCCAT CAAGCTACAG ATGGTCTTAC AAATGGAACC	780
CCAAATGAGC TCGACTAACA ACTTCTACTG AGGACCCCTG GACCGACCCA CTGGCCCTTT	840
AACTGGCTTA AAGAGTTTCC CTCTGGAGGA CACTACAAC TGCAGGGCCCC TTCTTTGCCC	900
CATCCACAGG AAGTTAGCTA GAGCAGTCAT CACCCAATTC CCAA	944

(2) INFORMATION FOR SEQ ID NO: 12: HE8

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 963 base pair
 (B) TYPE: nucleotide
 (C) STRANDS NUMBER: single
 (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TACAGGAACC CCATAATACG TCCTTGCAA ATTCTATTCA GCTCCAAC TGCTAGGAGTGG	60
CCCATTGTG CTGAACCCTC AAATCATGGG AATGAGAAAT GAATTTAGAC TGACCACAGC	120

CCTTATGAGT	TTTCAGCTAC	AGGGGTGTAT	AGAACCCTGA	TAAGGAGTTT	TCTTTGTGTG	180
TGGAAGATCC	TTCTATATTT	GCCTCCCCAC	CAACTGGACA	GGAAGTTGTA	CTTTAGCCTA	240
CATAGTACCT	CCTGTGACTT	ATCCTTTTCA	GAAGAGGCAG	TAGCTGTGCC	CATTCATGCT	300
AAGCTTCAGC	CGAGAGCAAT	CTCACTACTT	CCTCTATTGG	CTGGTTTAGG	ATTTACTACC	360
ACCTAGGAAG	TGGACTCACA	GCCTAGATGA	AATCTCTCTC	CAACTTACTC	AAATCCAGGA	420
CCAAATAGAC	TCATTAGCAG	CTGTGGTTCT	CCGAACCAGT	GAGCACTAGA	TCTCCAATCT	480
CCTCACTGCC	GAAAGGGGAG	GAACATGCCT	TTTTCTGAAC	AAGGAATGTT	GTTTTTATGT	540
CAATAAATCA	GGCATAGTGA	GAGATGGAAT	TAAATGACTT	CAGGATAGAG	CTAGCAGACT	600
ACATGGTGGG	ACAACCGAAA	CTACCTCAGG	GTTCTCACAG	CCTGTTCTCC	ACTGGCTTCT	660
TCCATTTTTA	GGTCCCTTCC	TTATGATTAT	TCTAGGAGTA	ACCTTTGGCC	CATGTCTTTT	720
CAGTTCCTTC	ATCCTTTCGT	TTCTTCCTGA	ATAGAATCAA	TGAAACTAGA	AATGTTACTG	780
CAGATGGAAC	CTCAGATGAC	TTCAACCAGC	ACCTATTATC	AAGGACCCCT	AAACCAGCCT	840
GCCGGCCCAT	ACCCGGACGT	TGACACCCAA	ACCACCTCTC	ACGAGGAAAC	CTCAGCTACA	900
GAACCCCTTC	TATGCCCTTA	TTCAGCAGGA	AGCAATTAGA	GTGGTCATCC	TCCCACACCC	960
CAA						963

(2) INFORMATION FOR SEQ ID NO: 13: HG8

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1362 base pair
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCACAATATC	CTCTTCAGG	AGGAGAACGA	TGGCCACCTG	AGGGAAGTAT	ACACTATAAT	60
ACCATCCTGC	AACTAGATCT	GTTTTGTAAA	CAAGAAGGCA	AGTGGATTTA	GGTACCATAT	120
G TTCAGACCT	TTTTCTCATT	AAGGGATGAT	AACCCACGAT	TGTGTAAGAC	ATGTAACCTG	180
CACCCACAG	GGAGTCCTCA	AATTCTACCC	CCATACCCAG	TCCTCCCCAC	GGCTCCTCCT	240
ACTAATGCCA	AACCCTCTCT	GGCTTCTACA	GCCAAAAGG	GAACAAATAA	AAGAGCCTTC	300
AGAGAGCCAA	GAGACCCAC	TGGCCCCTGG	CTATGTCCCTC	TTCAGGCTGT	AGGAGGGGAA	360
TTTGGCCCAA	CCCAGTACA	TGTTCCCTTT	TCTCTCTCTG	ATCTAAAGCA	AATTAAGGCA	420
GACTTGATG	AAAGTTCTCA	GATGACCCCA	ATAGATACGT	AGATGGCCTG	CTGGGTCTGG	480
GACAATCTTT	TGACCTTTCC	TGGAGAGAGA	TCATGTTATT	GCTTGATCAG	ACCTAACCTC	540

TAATGAGAAG AATGCTGCTT TAACAGGAGC CCGAGAGTTT GGGGATACCT GGTACCTCAG	600
TTAAGTAAGT GATAGAATGA CATCAGAAGA GAGCAGTTTC CTACTGGCCA GCAAGCAGTC	660
CCCAGTATGG ATCCCCACTG GGACCCTGAC TCGGATCATG GGGACTGGAG TCACAAACAT	720
TTACTGACCT GTATCCTAGA AGGGTTAAGG AGAACTAGGA AAAAGCCCAT GAACTATTCA	780
ATGATGTCTA CTATAACCCA AGGGAAGGAA GAAAACCCTA TTGCCTTCCT CAAAAGGCTG	840
AGGGAGGCTT TGAGAAAATA TACTCCCCTG TCACCAGATT CCCTCGAAGG CCAGTTAATT	900
TTAAAGGACA AATTTATTAC TCAGTCAGCT GCAGACATTA GGAAAAAGCT CCAAAGTTA	960
GCCTTGGGCC GAGCAAAATT TGGAGGCATC ATTAAACCTG GCAACCTCAG TGTTCTATCA	1020
TAGGGACCAA GAGGAACAGG CCGAAAAGGA AAAGCAGGAT AAGAGAAAGG CTGCAGATTT	1080
AGTCATGCCC TCAGACAAAC CTTGGCGGTT CAAAGAGGAG AAAAAATGGA GCAGGCCAAT	1140
CACCCAGCAG GGCTTATTAT CAGTGCAGTT TACAAGGACA CTTTAAACAA GATTGTCCAA	1200
AGAGAAATAA GCCGCCCTCT CACCCATGTC CACTATGCCA AGGTGATCAC TGGAAGGCAC	1260
ACTGTCCCAG AGGACAAAGG TTCTCTGGGC CAGAAGTCCC CAACCAGATG ATCCAGCAAC	1320
AGGATGGAGG GTGCCCCGGG CAAGCACCAG CTCGTGTTGT CA	1362

(2) INFORMATIONS FOR SEQ ID NO: 14: HE9

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 945 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 14:

TTGCAGATCA ATCTCAGACT GCTGTGCTAG CAATGAGTGA GGCTTCGTGG GCATGGGACC	60
CTCTGAGCCA GGCATGGGAT ATAATGTCCT TGTGTGCCAT TTGCTAAGAC TGTTGGAATA	120
GCACAGTATT AGGGTGGGAG TGGCCCGATT TTCCAGGTGC TGTCTGTCAC CGCTTCCCTT	180
GGCTAGGAAA GAGAATTCCC TGACCCCTTG TTCTTCCCAG GTAAGGCAGT GCCTCACCCCT	240
GCTTCAGCTC ACACTCAGGT GACTGCACCC ACTGTCCTGC CCCCCTGTC GGACAAGCCC	300
CAGTGAGATG AACCTGGTAC CTCAGTTGGA AATGCAGAAA TCACCTGTCT TCTGCGTCAC	360
TCACACTGGG AGCTGTAGAC TGGAGCTGTT CCTATTTGGC CATCTTGGA CCATCTCCCA	420
AATAGACTCT TTGGCAGCAG TGA CTCTCCA AAACCACCAA GGCCTAGACC TCCTCATTGC	480
TGAGAAAAGGA GGACTCTGCA CCTTCTTAGG GGAGGAGTGT TGTTTTTATA CTGACCAGTC	540
AGGGATGGTA CGAGATGCCA CCCGATGTTT ACAGGAAAAG GCTTCTGAAA TCACACAACA	600

CCTTTCAAAC TCTTATACCA ACCTCTGGAG TTGGGCAACA TGGCTTCTCC CCTTTCTCGG	660
TCCCATTGCA GCCATCTTGC TATTACTCGC CTTCAGGCTG TGTATTTTAA ACCTCCTTGT	720
CAAATTTGTT TCCTCTAGAA TTGAGGCCGT CAAGCTACAG ATGGTCTTAC AAATGGGACC	780
CCAAATGAGC TCAACTAACA ACTTCTGCCA AGGACCCCTG GACCAACCTG CTGGCCCTTT	840
CACTGGCCTT AAGAGTTCCC CTCTGGAGGG CACTACAACCT GCAGGGCCCC TTCTTTGCCC	900
CTATCCAGCA GGAAGTAGCT AGAGCAGTCA TCACCCAATT CCCAA	945

(2) INFORMATIONS FOR SEQ ID NO: 15: HE10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 939 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 15:

AGAGCTACCT TGGCAAGTAC TCTAGGAGTA TGGGAAAATG AAAACAACAA ACTCACACAC	60
CATTTTAACA TACACAATCA GGTCTGCCCA CCCAGCAAGG TATATTCTTT GTATGTGGAA	120
CATCGACCTA TATCTGCCTC CCCACTAACT AGACAGCCAC CTGAATCTTA GTCTTTCTAA	180
GTCCCAACAG TAACATTGCC CCAGGAAATC AGACCATATC AGTATCCCTC AAAGCTCAAG	240
TCTGTCAGTG CAGAGCCATA CAACTAATAC CCCTACTTAT AGGGTAAGGA ATGGCTACTG	300
CTACAGGAAC CAGAATAGCT AGTTTGTTTA CTTCAATTATC CTACTACCAC AACTCTCAA	360
ATGATTTCTC AGACAGTTTG CAAGAAATAA CGAAATCTAT CCTTACTCTA CAATCCCAA	420
TAGACTCCTT GGCAGCAGTG ACCCTCCAAA ACGGCTGAGG CCTAGACCTC CTCACTGCCA	480
AGAAAGGAGG ACTCTGCATT TTCTTAGGGG AAGAGTGTTT TTACACTAAC CAGTCAGGGA	540
CAGTATGAGA TGCCACTCGG AGTTTACAGG AAAAGGCTTC TGAAGTCAGA CAATGCCTTT	600
CAAACCTCTAT ACCAAACTCT GGAGTTGGGC AACATGGCTT CTCCCCTTTC TAGGTCCCGT	660
GACAGCCATC TTGCTATTAT TTGCCTTTGA GCCCTGTATT TTTAATCTCC TTTTCAAATT	720
TGTTTCCTCT GGATCGAGGC CATCGAGCTA CAGATGGTCT TCACAAATGG AACCCCAAAT	780
GAGCTCAACT AACAACTTCT ACTGAGGACC CCTGGACTAA CCTGCTGACC CTTTCACTGG	840
CCTGAAGAAT TCCCCTCTGG AGGACACTAC AACTGCAGGG CTCCTTCTTT GCCCCTATCC	900
AGCAGGAAGT AGCTAGAGCT GTCATTGCCT AATTCCTAA	939

(2) INFORMATIONS FOR SEQ ID NO: 16: HE11

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH : 979 base pairs

- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
AGTGATAATG GAATACTTGA AAGTAATCCC CTCACTCCCC AGGAACTAGT GCTCAGCTGG      60
CAGAACTAAT AGCCCTCACT CGGGTACTAG AATCAGGAGA AGGAAAAAGG GTAAATATAT      120
ATACAGACTC TAAGTGTGCT TACCTAGTCC TCCATGCCCA TGCAGCAATA TGGAGAGAAA      180
GGGAATTCCT AACTTCCGAG GGAACACCTA TCAAACATCA GGAAGCCATT AGGAAATTAT      240
TATTGGCTGT ACAGAAACCT AAAGAGGTGG CAGTTTACCA CTGCCGGGGT CATCAGAAAG      300
GAAAGGAAAAG GGAAATACAA GGGAGCCACC AAGTTGATAT TGAAGTCAAAG AGAGCCACAA      360
GGCTGGACCC TCCATTAGAA ATGCTTATAG GAGGACCCCT AGTATGGGGT AATCCCCTCC      420
GGGAAGCCAA GCCCCAGTAC TCAGCAGGAG AAATAGAATA GGGAAGTTCA TGAGGACATA      480
CTTCCCTCCC CTCCAGATGG CTAGCCACCA ATAAAGGAAA AATACTTTTG CCTGCAGCTA      540
ACCAATAGAA ATTACTTAAA ACCCTTCATC AAACCTTCCA CTTAGGCATT GATAGCACCC      600
ATGAGATGGC CAAATTATTA TTTACTGGAC CAGGCCTTTT CAAACTATC AAGCAGATAG      660
TCAGGGCCTG TAAAGTCTGC CAAAGAAATA ATCCCCTGCA CTGCAGGCCA TACATTTCAA      720
TCCCTGTATC TTTAACCTCC TTCTTAAATT TGTCTCTTCC AGAATCAAAG CTGTAAATT      780
ACAAATAGTT CTTCAAATGG AGCCACAGAT GCAGTCCATG ACTAAGATCC ACCACAGACC      840
CCTGGACCAG CCTGCTAGCC CATGCTCCAA TGTTAATGAC ATCGAAGGCA CCCCCTCCTG      900
AGGAAATCTC AACTGCACAA CCCCTACTAC GCCCCAATTC AGCAGAAAGC AGTTAGAGTG      960
GTCATCAGCC AACCTCCCC                                     979
```

(2) INFORMATION FOR SEQ ID NO: 17: HG11

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1774 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDS NUMBER: single
 - (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
CATGCTGGTAAAGGACCGCTAGAATCCAGCAGCCAGGACCACTTTCTTTGTGGTCAAGAAAGGTGGGAAAACAG
GTGCAGGACTGCTACACTGGTAAGCATAACTAATCCGATAAGCAGAGGTCCATGGGTGGTTACGCACCCTGGAAAGGAAT
AAGCATTAGGACTATAGAGGACACTCTAGGACTAATGCTCATCGGAAAATGACTAGGGGTACTGGCATCCCTATGTTCTT
TTTTCAGATGGGAAATGTTCCCCCAAGGCAGAAATGCCCTAAGATGTATTCTGGAGAAATGGGACCAATCTGACCATC
AGACACTAAGAAAGAAATGACTTATATTCTTCTGCAGTACCACCTGGCCACAATATCTTCTTCAAGGGGCAGAAACCTGG
```

CCTCCTGAGGGAAGTATAAATTATAACACCATCTTACAGCTAGACCTCTTTTGTAGAAAAGAAGGCCAAATGGAGTGAAGT
GCCATATGTACAAACTTTCTTTTCATTAAGAGATAACTCCCAATTATGTAAAAAGTGTGATTTATGCCCTACAGGAAGCC
CTCAGAGTCTACCTCCCGACCCAGCAAGACCCCAACTCCTTCTCCAATAATAAGGACCCCCCTTCAACCCAAATGGTC
CAAAAGGAGATAGACAAAGGGGTAAACAATGAACCAAAGAGTGCCAATATTACACGATTATACTCGCTCCAAGCAGTGGG
AGGAGAATTTGGCCCAGCCAGCGTGCATGTACCTTTTTCTCTCTCAGATTTAAAGCAAATTTAAATAGACCTAGGTAAAT
TCTCAGATAACCCCTGATGGCTATATTGATGTTTTACAAGGGTTAGGACAATCCTTTGATCTGACATGGAGAGATATAATG
TACTGCTAAATCAGACACTAACCCCAAATGAAAAAAGTGCTGCCATAACAGCAGCCTGAGAGTTTGGCGAACTCTGGTA
TCTCAGTCAGGTCAATGATAGGATGACAACAGATGAAAGAGAATGATTCCCCACAGGCCAGCAGGCAGTTCCCAGTGTAG
ACCCTCATTAGGACACAGAATCAGAACTTGGAGATTGGTGCCACAGACATTTGCTAACTTGCGTGCTAGAAGGACTAAGG
AAAAGTAGGAAGAAGCCCATGAATTATTCAATGATGTCCCCTATAACACAGGGAAAGGAAGAAAATCCTACTGCCTTTCT
GGAGAGACTAAGGGAAGGATTGAGGAAGCATACCTCCCTGTCACCTGACTCTATTAAAGGCCAACTAATCTTAAAGGATA
AGTTTATCACTCAGTCAGCTGCAGAGATTAAGAAAAAACTTCAAAGTATGCCTTAGGCCAGAGCAAACTTAGAAACC
CTACTGAACTTGGCAACCTCAGTTTTTTTATAATAGAGATCAGGAAGAGCAGGGGAATGGGACAAATGGGATAAAAAAAA
AAAAAAAGGTGACTGCTTTAGTCGTGGCCCTCAGGCAAATGGACTTTGGAGGCTCCAGAAAAGGGAAAAGCTGAGCAAAT
TGAATGCCTAACAGGGCTTGCTTCTAGTGTGGTCTACAAGGACACTTTAAAAAAGATTGTCCAAGTAGAAACAAGCTGCC
CCCTTGCTCCATGCCCTTATGTCAAGGGAATCACTGGAAGGCCACTGCCCCAGGAGATGAAGGTCTCTGAGTCAGAAG
CCACTAACCCAGATAATCCAGCAGCAGGACTGAGGATGCCCAGGGCAAGCGCCAGCCCATGCCATCACCCCTCACAGAGCCT
TGGGTATGCTTGACCATTGA

(2) INFORMATIONs for SEQ ID NO: 18: HE12

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 938 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TGTAGGAAGA ACTCCCTTCA GGACAGGACA ATAGATGGTT CCTCCCAGGT GATTAAGGAA	60
AAAAGACACA GTATTCAGTA AGTGATAAGG AAACTCTTGT AGAAGCAGAG TTAGAAAAAT	120
TGCCTAATAA TTGGTCTGCT CAAATGTGTG AGTTGTTTGC ACTCAGCCAA ATCTTAAAGT	180
ACTTACAGAA TCAGGAAGCA GCCATCTATA CCAATTCTAA GTTAATATGG ACTAAACGAG	240
GTTTTATTAG TAGCAAAGAA AAATTAAAAT CCCAACTTA CAAGGTTTTT CAACTAAAGTT	300
TGCCAAAAGT TAACAGTGTA ACATGTATTA TCCTACTATC ACACACTCTC AAAGGATTTT	360
TCAGACAGTT TGCAAGAAAT AACGTAATCT ATCCTTACTC TACAGTCCCA AATAGACTCT	420
TTGGTAGCAG TGA CTCTCCA AAACTGCCGA GGTCTAGACC TCCTCAATGC TGAGAAAGGA	480
GAACTCTGCA CCTTCTTAGG GGAAGAGTGC TGTTTTTTACA CTAACCAGTC AGGGATAGTA	540
TGAGATACTG CCTGACGTTT ACAGGAAAAG GCTTCTGAAA TCAGACAACG CCTTTCAAGC	600
TCTTATACCA ACCTCTGGAG TTGGGCAACA TGGCTTCTCC CCTTGCTAGG TCCTGTGGCA	660
GCCATCTTGC TATTACTTGC CTTGGGGCCC TGTATTTTTA ACCTCCTTGT CAAATTTGTT	720
TCCTCTAGGA TCAAGGCCAT CAAGCTACAG ATGGTCTTAC AAATGGAACC CCAAATGAGC	780
TCAACTAACA ACTTCTACTG AGGACACCTG GACTGACCCA CTGGCCCTTT CACTGGCCTA	840
AAGAGTTCCC TTCTGGAGGA CACTACAAC TGCAGGGCCCC GTCTTCACCC CTATCCAGCA	900

GGAAGTAGCT AGATCAGTCA TTGCCCAATT CCCAACAG

938

(2) INFORMATIONS FOR SEQ ID NO: 19: HG12

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1308 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GATGCTTGCC CCAGGCACCC TCAGTCCTGT TGTGGATCA TCTGGTCGGG GGCTTCTGGC	60
CCAAAGAACC TTTGTCCTCT GAGGCAGTGC ACCTTCCAGT GATTGCCTCA GCATTGTGGA	120
CATGGGCAAG GGGGCAGCTT GTTTCTCACT GGACAATCTT TTTTAAGGTG TCCTTCCAAA	180
CCACACTGGT AACAAGCCCT ACCAGGTGAT TGGCCTGCTC TATTTTCTGT CCTCTCTGAA	240
CCACCAAGGT TTGTCTGTCT GAGGGTCATG ACTAAGGCTG TGGCCTTTCT CTGATCTTGC	300
TTTTCTTTT TGGCCTGTTC CTCTTGGTAC CTATTATAGA AACTGAGGT TGCCAGGTTT	360
AACAATGGCT CCAGATTTTG TTCAGGGCAC AGGGCTCATT TTGGAGCTTT CTCCTGATAT	420
CTGCAGCTGA TTGGGTAATA AACTTATCTT TTAGGATCAA TTGACTCTCA AGAGAGTTGG	480
GTGACAGGGG AGTATATTTT CTTGAGGCCT CCCATAGCCG CTCTAGGAAG GCAGAAGGAT	540
TTTCTTCCTT TCCCTGAGTT ATAAAAGACA TCATTGAACA ACTCATGGAC TTTTTCCTT	600
TTCTCCGTAG TCCTTCTAGA ACACAGGTCA GCAGATGTTT ACGACTCCAG TCCCCATGAT	660
CTGAGTCTAG ACACCAAGTG GGATCCATAC TGGGGATGGC CTGCTGACTG GTAGGGAATT	720
TGTCCCTTTC TTTGGCTGTC ATTCTATCAT TTAATTGACT AAGATACCAA GTATCTCCAA	780
ATTCTCAGGC TGCAGCTAAA GCTGCATTCT TTTTATTAAA GGCCAGGGTT TGATCTAATA	840
GCATGACATC TCTCCAAGTG AGGTCAAAGG TTTGCCCTAG ATCCATAGGA CATCAGAGAA	900
GGAGAAGGGG ACATACACCT GAGTTAGCCA AATTCCCCTC CCTCTACAGC TTGAAGGGGA	960
CATAAGCAAT AGCCTGGGGA TTTTGTGGT CCTTTGGAGA TTTCTTTGCT TGTTTCCTTC	1020
TGGGTGGGGG AGATTAGAGG AGGCTTATCA GTAATAGGAA GGGGAGCTAT AGGGAGGCTA	1080
GGATATGGGG GTAAGCTGAG AGGTCATCTT GTGGGATGTA AATTGCAAGC TTTGCATAGT	1140
TGTGGATTTT CCTTACAATG AAAATAAAGC TTGGACATAA GGTATTTTAC TCCATTTGCC	1200
TTCCCTCTTA CAGAAAAGGT CAAGCTGCAG GATAGTACTG TAATTTATAC TTCCTTCAGG	1260
TGGCCATTTT TTCCCATCAG AGAGAGAATA CTGGGGCTGG GCCATAGT	1308

(2) INFORMATIONS FOR SEQ ID NO: 20: R1

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 711 base pairs
 (B) TYPE: nucleotide
 (C) STRANDS NUMBER: single
 (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACTGAGAGAC AGGACTAGCT GGATTTCTTA GGCCGACTAA GAATCCCTAA GCCTAGCTGG	60
GAAGGTGACC ACGTCCACCT TTAAACACGG GGCTTGCAAC TTAGCTCACA CCTGACCAAT	120
CAGAGAGCTC ACTAAAATGC TAATTAGGCA AAGACAGGAG GTAAAGAAAT AGCCAATCAT	180
CTATTGCCTG AGAGCACAGC AGGAGGGACA ACAATCGGGA TATAAACCCA GGCATTCGAG	240
CTGGCAACAG CAGCCCCCTT TTGGGTCCCT TCCCTTTGTA TGGGAGCTGT TTTCATGCTA	300
TTTCACTCTA TTAAATCTTG CAACTGCACT CTTCTGGTCC ATGTTTCTTA CGGCTCGAGC	360
TGAGCTTTTG CTCACCGTCC ACCACTGCTG TTTGCCACCA CCGCAGACCT GCCGCTGACT	420
CCCATCCCTC TGGATCCTGC AGGGTGTCCG CTGTGCTCCT GATCCAGCGA GGCGCCCATT	480
GCCGCTCCCA ATTGGGCTAA AGGCTTGCCA TTGTTCTGCT ACGGCTAAGT GCCTGGGTTT	540
GTTCTAATTG AGCTGAACAC TAGTCACTGG GTTCCATGGT TCTCTTCTGT GACCCACGGC	600
TTCTAATAGA ACTATAACAC TTACCACATG GCCCAAGATT CCATTCCTTG GAATCCGTGA	660
GGCCAAGAAC TCCAGGTCAG AGAATACGAG GCTTGCCACC ATCTTGGAAG C	711

(2) INFORMATION FOR SEQ ID NO: 21: R1F

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 711 base pairs
 (B) TYPE: nucleotide
 (C) STRANDS NUMBER: single
 (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ACTGAGAGAC AGGACTAGCT GGATTTCTTA GGCTGACTAA GAATCCCTAA GCCTAGCTGG	60
GAAGGTGACC ACATCCACCT TTAAACACGG GGCTTGCAAC TTAGCTCACA CCTGACCAAT	120
CAGAGAGCTC ACTAAAATGC TAATTAGGCA AAGACAGGAG GTAAAGAAAT AGCCAATCAT	180
CTATTGCCTG AGAGCACAGC AGGAGGGACA ATGATCGGGA TATAAACCCA AGTCTTCGAG	240
CCGGCAACGG CAACCCCCTT TGGGTCCCCT CCCTTTGTAT GGGAGCTCTG TTTTCATGCT	300
ATTTCACTCT ATTAAATCTT GCAACTGCAC TCTTCTGGTC CATGTTTCTT ACGGCTTGAG	360

```

CTGAGCTTTC GCTCGCCATC CACCACTGCT GTTTGCCGCC ACCGCAGACC CGCCGCTGAC      420
TCCCATCCCT CTGGATCATG CAGGGTGTCC GCTGTGCTCC TGATCCAGCG AGGCACCCAT      480
TGCCGCTCCC AATCGGGCTA AAGGCTTGCC ATTGTTCTTG CATGGCTAAG TGCCTGGGTT      540
CATCCTAATT GAGCTGAACA CTAGTCACTG GGTTCATGG TTCTCTTCTG TGACCCACAG      600
CTTCTAATAG AGCTATAACA CTCACCGCAT GGCCCAAGGT TCCATTCCTT GAATCCATAA      660
GGCCAAGAAC CCCAGGTCAG AGAACACGAG GCTTGCCACC ATCTTGGGAG C                711

```

(2) INFORMATIONS FOR SEQ ID NO: 22: HERV-7q (CODING SEQUENCE WITH 3 READING FRAMES)

(i) SEQUENCE CHARACTERISTICS:
 (B) TYPE: nucleotide
 (C) STRANDS NUMBER: single
 (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```

AAGCTCCTTCAGGAGAACAAGAACAGGCCATTACCCTGGAGAAGACTGGCAACTGATTTTACCCACAAGCCCAA
LysLeuLeuGlnGluAsnLysGluGlnAlaIleThrLeuGluLysThrGlyAsn...PheTyrProGlnAlaGln
SerSerPheArgArgThrLysAsnArgProLeuProTrpArgArgLeuAlaThrAspPheThrHisLysProLys
AlaProSerGlyGluGlnArgThrGlyHisTyrProGlyGluAspTrpGlnLeuIleLeuProThrSerProAsn

ACCTCAGGGATTTTCACTATCTACTAGTCTGGGTAGATACTTTCACGGGTTGGGCAGAGGCCTTCCCCTGTAGGAC
ThrSerGlyIleSerValSerThrSerLeuGlyArgTyrPheHisGlyLeuGlyArgGlyLeuProLeu...Asp
ProGlnGlyPheGlnTyrLeuLeuValTrpValAspThrPheThrGlyTrpAlaGluAlaPheProCysArgThr
LeuArgAspPheSerIleTyr...SerGly...IleLeuSerArgValGlyGlnArgProSerProValGlyGln

AGAAAAGGCCCAAGAGGTAATAAAGGCACTAGTTCATGAAATAATTCCCAGATTTCGGACTTCCCCGAGGCTTACA
ArgLysGlyProArgGlyAsnLysGlyThrSerSer...AsnAsnSerGlnIleArgThrSerProArgLeuThr
GluLysAlaGlnGluValIleLysAlaLeuValHisGluIleIleProArgPheGlyLeuProArgGlyLeuGln
LysArgProLysArg.....ArgHis...PheMETLys...PheProAspSerAspPheProGluAlaTyrArg

GAGTGACAATAGCCCTGCTTTCCAGGCCACAGTAACCCAGGGAGTATCCAGGCGTTAGGTATACGATATCACTT
Glu...Gln...ProCysPheProGlyHisSerAsnProGlySerIleProGlyValArgTyrThrIleSerLeu
SerAspAsnSerProAlaPheGlnAlaThrValThrGlnGlyValSerGlnAlaLeuGlyIleArgTyrHisLeu
ValThrIleAlaLeuLeuSerArgProGln...ProArgGluTyrProArgArg...ValTyrAspIleThrTyr

ACACTGCGCCTGAAGGCCACAGTCCTCAGGGAAGGTCGAGAAAAATGAATGAAACACTCAAAGGACATCTAAAAAA
ThrLeuArgLeuLysAlaThrValLeuArgGluGlyArgGluAsnGlu...AsnThrGlnArgThrSerLysLys
HisCysAla...ArgProGlnSerSerGlyLysValGluLysMETAsnGluThrLeuLysGlyHisLeuLysLys
ThrAlaProGluGlyHisSerProGlnGlyArgSerArgLys...METLysHisSerLysAspIle...LysSer

GCAAACCCAGGAAACCCACCTCACATGGCCTGCTCTGTTGCCTATAGCCTTAAAAAGAATCTGCAACTTTCCCCA
385      395      405      415      425      435      445
AlaAsnProGlyAsnProProHisMETAlaCysSerValAlaTyrSerLeuLysLysAsnLeuGlnLeuSerPro
GlnThrGlnGluThrHisLeuThrTrpProAlaLeuLeuProIleAlaLeuLysArgIleCysAsnPheProGln
LysProArgLysProThrSerHisGlyLeuLeuCysCysLeu...Pro...LysGluSerAlaThrPheProLys

AAAAGCAGGACTTAGCCCATACGAAATGCTGTATGGAAGGCCCTTCATAACCAATGACCTTGTGCTTGACCCAAG
LysSerArgThr...ProIleArgAsnAlaValTrpLysAlaLeuHisAsnGln...ProCysAla...ProLys
LysAlaGlyLeuSerProTyrGluMETLeuTyrGlyArgProPheIleThrAsnAspLeuValLeuAspProArg
LysGlnAspLeuAlaHisThrLysCysCysMETGluGlyProSer...ProMETThrLeuCysLeuThrGlnAsp

```

ACAGCCAACTTAGTTGCAGACATCACCTCCTTAGCCAAATATCAACAAGTTCTTAAACATTACAAGGAACCTAT
ThrAlaAsnLeuValAlaAspIleThrSerLeuAlaLysTyrGlnGlnValLeuLysThrLeuGlnGlyThrTyr
GlnProThr...LeuGlnThrSerProPro...ProAsnIleAsnLysPheLeuLysHisTyrLysGluProIle
SerGlnLeuSerCysArgHisHisLeuLeuSerGlnIleSerThrSerSer...AsnIleThrArgAsnLeuSer

CCCTGAGAAGAGGGAAAAGAACTATTCCACCCTTGTGACATGGTATTAGTCAAGTCCCTTCCCTCTAATTCCCCA
Pro...GluGluGlyLysGluLeuPheHisProCysAspMETValLeuValLysSerLeuProSerAsnSerPro
ProGluLysArgGluLysAsnTyrSerThrLeuValThrTrpTyr...SerSerProPheProLeuIleProHis
LeuArgArgGlyLysArgThrIleProProLeu...HisGlyIleSerGlnValProSerLeu...PheProIle

TCCCTAGATACATCCTGGGAAGGACCCTACCCAGTCATTTTATCTACCCCAACTGCGGTTAAAGTGGCTGGAGTG
SerLeuAspThrSerTrpGluGlyProTyrProValIleLeuSerThrProThrAlaValLysValAlaGlyVal
Pro...IleHisProGlyLysAspProThrGlnSerPheTyrLeuProGlnLeuArgLeuLysTrpLeuGluTrp
ProArgTyrIleLeuGlyArgThrLeuProSerHisPheIleTyrProAsnCysGly...SerGlyTrpSerGly

GAGTCTTGGATACATCACACTTGAGTCAAATCCTGGATACTGCCAAAGGAACCTGAAAATCCAGGAGACAACGCT
GluSerTrpIleHisHisThr...ValLysSerTrpIleLeuProLysGluProGluAsnProGlyAspAsnAla
SerLeuGlyTyrIleThrLeuGluSerAsnProGlyTyrCysGlnArgAsnLeuLysIleGlnGluThrThrLeu
ValLeuAspThrSerHisLeuSerGlnIleLeuAspThrAlaLysGlyThr...LysSerArgArgGlnArg...

AGCTATTCTGTGAACCTCTAGAGGATTTGCGCCTGCTCTTCAAACAACAACCAGGAGGAAAGTAACTAAAATCA
SerTyrSerCysGluProLeuGluAspLeuArgLeuLeuPheLysGlnGlnProGlyGlyLys...LeuLysSer
AlaIleProValAsnLeu...ArgIleCysAlaCysSerSerAsnAsnAsnGlnGluGluSerAsn...AsnHis
LeuPheLeu...ThrSerArgGlyPheAlaProAlaLeuGlnThrThrThrArgArgLysValThrLysIleIle

TAAATCCCCATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACTGCACCC
...IleProMETAlaLeuProTyrHisIlePheLeuPheThrValLeuLeuProSerPheThrLeuThrAlaPro
LysSerProTrpProSerLeuIleIlePhePheSerLeuLeuPhePheTyrProLeuSerLeuSerLeuHisPro
AsnProHisGlyProProLeuSerTyrPheSerLeuTyrCysSerPheThrLeuPheHisSerHisCysThrPro

CCTCCATGCCGCTGTATGACCAGTAGCTCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATT
ProProCysArgCysMETThrSerSerSerProTyrGlnGluPheLeuTrpArgMETGlnArgProGlyAsnIle
LeuHisAlaAlaVal...ProValAlaProLeuThrLysSerPheTyrGlyGluCysSerValProGluIleLeu
SerMETProLeuTyrAspGln...LeuProLeuProArgValSerMETGluAsnAlaAlaSerArgLysTyr...

GATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCACACCCATATGCCCGCAACTGC
AspAlaProSerTyrArgSerLeuSerLysGlyThrProThrPheThrAlaHisThrHisMETProArgAsnCys
METProHisArgIleGlyValPheLeuArgGluProProProSerLeuProThrProIleCysProAlaThrAla
CysProIleVal...GluSerPhe...GlyAsnProHisLeuHisCysProHisProTyrAlaProGlnLeuLeu

TATCACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCT
TyrHisSerAlaThrLeuCysMETHisAlaAsnThrHisTyrTrpThrGlyLysMETIleAsnProSerCysPro
IleThrLeuProLeuPheAlaCysMETGlnIleLeuIleIleGlyGlnGluLys...LeuIleLeuValValLeu
SerLeuCysHisSerLeuHisAlaCysLysTyrSerLeuLeuAspArgLysAsnAsp...Ser...LeuSerTrp

GGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCAACTGGTATGTCTGATGGGGGTGGAGTTCAAGAT
GlyGlyLeuGlyValThrValCysTrpThrTyrPheThrGlnThrGlyMETSerAspGlyGlyGlyValGlnAsp
GluAspLeuGluSerLeuSerValGlyLeuThrSerProLysLeuValCysLeuMETGlyValGluPheLysIle
ArgThrTrpSerHisCysLeuLeuAspLeuLeuHisProAsnTrpTyrVal...TrpGlyTrpSerSerArgSer

CAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCCTACAAA
GlnAlaArgGluLysHisValLysGluValIleSerGlnLeuThrArgValHisGlyThrSerSerProTyrLys
ArgGlnGluLysAsnMET...LysLys...SerProAsnSerProGlyTyrMETAlaProLeuAlaProThrLys
GlyLysArgLysThrCysLysArgSerAsnLeuProThrHisProGlyThrTrpHisLeu...ProLeuGlnArg

GGACTAGATCTCTCAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCTC
GlyLeuAspLeuSerLysLeuHisGluThrLeuArgThrHisThrArgLeuValSerLeuPheAsnThrThrLeu
Asp...IleSerGlnAsnTyrMETLysProSerValProIleLeuAlaTrp...AlaTyrLeuIleProProSer
ThrArgSerLeuLysThrThr...AsnProProTyrProTyrSerProGlyLysProIle...TyrHisProHis

ACTGGGCTCCATGAGGTCTCGGCCCAAACCCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATAT
ThrGlyLeuHisGluValSerAlaGlnAsnProThrAsnCysTrpIleCysLeuProLeuAsnPheArgProTyr
LeuGlySerMETArgSerArgProLysThrLeuLeuThrValGlyTyrAlaSerPro...ThrSerGlyHisMET
TrpAlaPro...GlyLeuGlyProLysProTyr...LeuLeuAspMETProProProGluLeuGlnAlaIleCys

GTTTCAATCCCTGTACCTGAACAATGGAACAACCTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCT
ValSerIleProValProGluGlnTrpAsnAsnPheSerThrGluIleAsnThrThrSerValLeuValGlyPro
PheGlnSerLeuTyrLeuAsnAsnGlyThrThrSerAlaGlnLys...ThrProLeuProPhe.....AspLeu
PheAsnProCysThr...ThrMETGluGlnLeuGlnHisArgAsnLysHisHisPheArgPheSerArgThrSer

CTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACC
LeuValSerAsnLeuGluIleThrHisThrSerAsnLeuThrCysValLysPheSerAsnThrThrTyrThrThr
LeuPheProIleTrpLys...ProIleProGlnThrSerProVal...AsnLeuAlaIleLeuHisThrGlnPro
CysPheGlnSerGlyAsnAsnProTyrLeuLysProHisLeuCysLysIle...GlnTyrTyrIleHisAsnGln

AACTCCCAATGCATCAGGTGGGTAACCTCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGT
AsnSerGlnCysIleArgTrpValThrProProThrGlnIleValCysLeuProSerGlyIlePhePheValCys
ThrProAsnAlaSerGlyGly...LeuLeuProHisLys...SerAlaTyrProGlnGluTyrPheLeuSerVal
LeuProMETHisGlnValGlyAsnSerSerHisThrAsnSerLeuProThrLeuArgAsnIlePheCysLeuTrp

GGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATG
GlyThrSerAlaTyrArgCysLeuAsnGlySerSerGluSerMETCysPheLeuSerPheLeuValProProMET
ValProGlnProIleValVal...METAlaLeuGlnAsnLeuCysAlaSerSerHisSer...CysProLeu...
TyrLeuSerLeuSerLeuPheGluTrpLeuPheArgIleTyrValLeuProLeuIleLeuSerAlaProTyrAsp

ACCATCTACACTGAACAAGATTTATACAGTTATGTGCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCT
ThrIleTyrThrGluGlnAspLeuTyrSerTyrValIleSerLysProArgAsnLysArgValProIleLeuPro
ProSerThrLeuAsnLysIleTyrThrValMETSerTyrLeuSerProAlaThrLysGluTyrProPhePheLeu
HisLeuHis...ThrArgPheIleGlnLeuCysHisIle...AlaProGlnGlnLysSerThrHisSerSerPhe

TTTGTATATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACTCTACTCAGTTCTAC
PheValIleGlyAlaGlyValLeuGlyAlaLeuGlyThrGlyIleGlyGlyIleThrThrSerThrGlnPheTyr
LeuLeu...GluGlnGluCys...ValHis...ValLeuAlaLeuAlaValSerGlnProLeuLeuSerSerThr
CysTyrArgSerArgSerAlaArgCysThrArgTyrTrpHisTrpArgTyrHisAsnLeuTyrSerValLeuLeu

TACAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTGCGCGACTCCCTGGTCACCTTGCAAGATCAACTT
TyrLysLeuSerGlnGluLeuAsnGlyAspMETGluArgValAlaAspSerLeuValThrLeuGlnAspGlnLeu
ThrAsnTyrLeuLysAsn...METGlyThrTrpAsnGlySerProThrProTrpSerProCysLysIleAsnLeu
GlnThrIleSerArgThrLysTrpGlyHisGlyThrGlyArgArgLeuProGlyHisLeuAlaArgSerThr...

AACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGT
AsnSerLeuAlaAlaValValLeuGlnAsnArgArgAlaLeuAspLeuLeuThrAlaGluArgGlyGlyThrCys
ThrPro...GlnGln...SerPheLysIleGluGluLeu...ThrCys...ProLeuLysGluGlyGluProVal
LeuProSerSerSerSerProSerLysSerLysSerPheArgLeuAlaAsnArg...LysArgGlyAsnLeuPhe

TTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGA
LeuPheLeuGlyGluGluCysCysTyrTyrValAsnGlnSerGlyIleValThrGluLysValLysGluIleArg
TyrPhe...GlyLysAsnAlaValIleMETLeuIleAsnProGluSerSerLeuArgLysLeuLysLysPheGlu
IlePheArgGlyArgMETLeuLeuLeuCys...SerIleArgAsnArgHis...GluSer...ArgAsnSerArg

GATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGG
AspArgIleGlnArgArgAlaGluGluLeuArgAsnThrGlyProTrpGlyLeuLeuSerGlnTrpMETProTrp
IleGluTyrAsnValGluGlnArgSerPheGluThrLeuAspProGlyAlaSerSerAlaAsnGlyCysProGly
SerAsnThrThr...SerArgGlyAlaSerLysHisTrpThrLeuGlyProProGlnProMETAspAlaLeuAsp

ATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTT
IleLeuProPheLeuGlyProLeuAlaAlaIleIleLeuLeuLeuLeuPheGlyProCysIlePheAsnLeuLeu

PheSerProSer...AspLeu...GlnLeu...TyrCysTyrSerSerLeuAspProValSerLeuThrSerLeu
SerProLeuLeuArgThrSerSerSerTyrAsnIleAlaThrProLeuTrpThrLeuTyrLeu...ProProCys

GTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATC
ValAsnPheValSerSerArgIleGluAlaValLysLeuGlnMETGluProLysMETGlnSerLysThrLysIle
LeuThrLeuSerLeuProGluSerLysLeu...AsnTyrLysTrpSerProArgCysSerProArgLeuArgSer
...LeuCysLeuPheGlnAsnArgSerCysLysThrThrAsnGlyAlaGlnAspAlaValGlnAsp...AspLeu

TACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAA
TyrArgArgProLeuAspArgProAlaSerProArgSerAspValAsnAspIleLysGlyThrProProGluGlu
ThrAlaAspProTrpThrGlyLeuAlaHisAspLeuMETLeuMETThrSerLysAlaProLeuLeuArgLys
ProGlnThrProGlyProAlaCys...ProThrIle...Cys.....HisGlnArgHisProSer...GlyAsn

ATCTCAGCTGCACAACCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCTGCGCCAACCTCCCCA
IleSerAlaAlaGlnProLeuLeuArgProAsnSerAlaGlySerSer...SerGlyArgArgProThrSerPro
SerGlnLeuHisAsnLeuTyrTyrAlaProIleGlnGlnGluAlaValArgAlaValValGlyGlnProProGln
LeuSerCysThrThrSerThrThrProGlnPheSerArgLysGlnLeuGluArgSerSerAlaAsnLeuProAsn

ACAGCACTTAGGTTTTCTGTTGAGATGGGGG
ThrAlaLeuArgPheSerCys...AspGlyGly
GlnHisLeuGlyPheProValGluMETGly
SerThr...ValPheLeuLeuArgTrpGly

(2) INFORMATIONS FOR SEQ ID NO: 23: HERV-7q (DEDUCED ENV PROTEINS)

(i) SEQUENCE CHARACTERISTICS:

- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

P K T A N L V A D I T S L A K Y Q Q V L K T L Q G
CCCCAAGACAGCCAACCTTAGTTGCAGACATCACCTCCTTAGCCAAATATCAACAAGTTCTTAAAACATTACAAGGA
T Y P X E E G K E L F H P C D M V L V K S L P S N
ACCTATCCCTGAGAAGAGGGAAAAGAACTATTCCACCCCTGTGACATGGTATTAGTCAAGTCCCTTCCCTCTAAT
S P S L D T S W E G P Y P V I L S T P T A V K V A
TCCCCATCCCTAGATACATCCTGGGAAGGACCCTACCCAGTCATTTTATCTACCCCAACTGCGGTTAAAGTGGCT
G V E S W I H H T X V K S W I L P K E P E N P G D
GGAGTGGAGTCTTGGATACATCACACTTGAGTCAAATCCTGGATACTGCCAAAGGAACCTGAAAATCCAGGAGAC
N A S Y S C E P L E D L R L L F K Q Q P G G K * L
AACGCTAGCTATTCTGTGAACCTCTAGAGGATTTGCGCCTGCTCTTCAAACAACAACCAGGAGGAAAGTAACTA
K S X I P M A L P Y H I F L F T V L L P S F T L T
AAATCATAAATCCCCATGGCCCTCCCTTATCATATTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT
A P P P C R C M T S S S P Y Q E F L W R M Q R P G
GCACCCCTCCATGCCGCTGTATGACCAGTAGCTCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGTCCCGGA
N I D A P S Y R S L S K G T P T F T A H T H M P R
AATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCGC
N C Y H S A T L C M H A N T H Y W T G K M I N P S
AACTGCTATCACTTGCCACTCTTTGCATGCAATGCAATCTATTGGACAGGAAAATGATTAATCCTAGT
C P G G L G V T V C W T Y F T Q T G M S D G G G V
TGTCCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT
Q D Q A R E K H V K E V I S Q L T R V H G T S S P
CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTACCCGGGTACATGGCACCTCTAGCCCC
Y K G L D L S K L H E T L R T H T R L V S L F N T

TACAAAGGACTAGATCTCTCAAACCTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACC
T L T G L H E V S A Q N P T N C W I C L P L N F R
ACCCTCACTGGGCTCCATGAGGTCTCGGCCAAAAACCCTACTAAGTGTGGATATGCCTCCCCCTGAACTTCAGG
P Y V S I P V P E Q W N N F S T E I N T T S V L V
CCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACCTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA
G P L V S N L E I T H T S N L T C V K F S N T T Y
GGACCTCTTGTTCCTCAATCTGGAATAAACCATAACCTCAAACCTCACCTGTGTAAATTTAGCAATACTACATAC
T T N S Q C I R W V T P P T Q I V C L P S G I F F
ACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTT
V C G T S A Y R C L N G S S E S M C F L S P
GTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAACTATGTGCTTCCTCTCATTTAGTGCCC
P M T I Y T E Q D L Y S Y V I S K P R N K R V P I
CCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTGCATATCTAAGCCCCGCAACAAAAGAGTACCCATT
L P F V I G A G V L G A L G T G I G G I T T S T Q
CTTCTTTTGTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAG
F Y Y K L S Q E L N G D M E R V A D S L V T L Q D
TTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGAT
Q L N S L A A V V L Q N R R A L D L L T A E R G G
CAACTTAACCTCCCTAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGA
T C L F L G E E C C Y Y V N Q S G I V T E K V K E
ACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAAGTTAAAGAA
I R D R I Q R R A E E L R N T G P W G L L S Q W M
ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATG
P W I L P F L G P L A A I I L L L L F G P C I F N
CCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAAC
L L V N F V S S R I E A V K L Q M E P K M Q S K T
CTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACT
K I Y R R P L D R P A S P R S D V N D I K G T P P
AAGATCTACCGCAGACCCCTGGACCGGCCTAGTACCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCT
E E I S A A Q P L L R P N S A G S S X S G R R P T
GAGGAAATCTCAGCTGCACAACCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCTGCGCCAACC
S P T A L R F S C X
TCCCCAACAGCACTTAGGTTTTCTGTTGA

(2) INFORMATION FOR SEQ ID NO: 24: HERV-7q (GAG CODING SEQUENCE)

(i) SEQUENCE CHARACTERISTICS:

- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

T	S	F	V	E	K	A	N	G	V	K	C	H	K	Y
ACC	TCT	TTT	GTA	GAA	AAG	GCA	AAT	GGA	GTG	AAG	TGC	CAT	AAG	TAC
K	L	S	F	H	X	E	T	T	H	N	Y	V	K	S
AAA	CTT	TCT	TTT	CAT	TAA	GAG	ACA	ACT	CAC	AAT	TAT	GTA	AAA	AGT
V	I	Y	A	L	Q	E	A	F	R	V	Y	L	P	I
GTG	ATT	TAT	GCC	CTA	CAG	GAA	GCC	TTC	AGA	GTC	TAC	CTC	CCT	ATC
P	A	S	P	T	P	S	P	T	N	K	D	P	P	S
CCA	GCA	TCC	CCG	ACT	CCT	TCC	CCA	ACT	AAT	AAG	GAC	CCC	CCT	TCA
T	Q	M	V	Q	K	E	I	D	K	R	V	N	S	E
ACC	CAA	ATG	GTC	CAA	AAG	GAG	ATA	GAC	AAA	AGG	GTA	AAC	AGT	GAA
P	K	S	A	N	I	P	Q	L	X	P	L	Q	A	V
CCA	AAG	AGT	GCC	AAT	ATT	CCC	CAA	TTA	TGA	CCC	CTC	CAA	GCA	GTG
G	G	R	E	F	G	P	A	R	V	H	V	P	F	S
GGA	GGA	AGA	GAA	TTC	GGC	CCA	GCC	AGA	GTG	CAT	GTG	CCT	TTT	TCT

```

L   P   D   L   K   Q   I   K   T   D   L   G   K   F   S
CTC CCA GAC TTA AAG CAA ATA AAA ACA GAC TTA GGT AAA TTC TCA
D   N   P   D   G   Y   I   D   V   L   Q   G   L   G   Q
GAT AAC CCT GAT GGC TAT ATT GAT GTT TTA CAA GGG TTA GGA CAA
F   F   D   L   T   W   R   D   I   M   S   L   L   N   Q
TTC TTT GAT CTG ACA TGG AGA GAT ATA ATG TCA CTG CTA AAT CAG
T   L   T   P   N   E   R   S   A   T   I   T   A   A   X
ACA CTA ACC CCA AAT GAG AGA AGT GCC ACC ATA ACT GCA GCC TGA
E   F   G   D   L   W   Y   L   S   Q   V   N   D   R   M
GAG TTT GGC GAT CTC TGG TAT CTC AGT CAG GTC AAT GAT AGG ATG
T   T   E   E   R   E   X   F   P   T   G   Q   Q   A   V
ACA ACA GAG GAA AGA GAA TGA TTC CCC ACA GGC CAG CAG GCA GTT
P   S   L   D   P   H   W   D   T   E   S   E   H   G   D
CCC AGT CTA GAC CCT CAT TGG GAC ACA GAA TCA GAA CAT GGA GAT
W   C   C   R   H   L   L   T   C   V   L   E   G   L   R
TGG TGC TGC AGA CAT TTG CTA ACT TGT GTG CTA GAA GGA CTA AGG
K   T   R   K   K   S   M   N   Y   S   M   M   S   T   I
AAA ACT AGG AAG AAG TCT ATG AAT TAC TCA ATG ATG TCC ACC ATA
T   Q   G   R   E   E   N   P   T   A   F   L   E   R   L
ACA CAG GGA AGG GAA GAA AAT CCT ACT GCC TTT CTG GAG AGA CTA
R   E   A   L   R   K   R   A   S   L   S   P   D   S   S
AGG GAG GCA TTG AGG AAG CGT GCC TCT CTG TCA CCT GAC TCT TCT
E   G   Q   L   I   L   K   R   K   F   I   T   Q   S   A
GAA GGC CAA CTA ATC TTA AAG CGT AAG TTT ATC ACT CAG TCA GCT
A   D   I   R   K   K   L   Q   K   S   A   V   G   P   E
GCA GAC ATT AGA AAA AAA CTT CAA AAG TCT GCC GTA GGC CCG GAG
Q   N   L   E   T   L   L   N   L   A   T   S   V   F   Y
CAA AAC TTA GAA ACC CTA TTG AAC TTG GCA ACC TCG GTT TTT TAT
N   R   D   Q   E   E   Q   A   E   Q   D   K   R   D   X
AAT AGA GAT CAG GAG GAG CAG GCG GAA CAG GAC AAA CGG GAT TAA
K   K   G   H   R   F   S   H   D   P   Q   A   S   G   L
AAA AAA GGC CAC CGC TTT AGT CAT GAC CCT CAG GCA AGT GGA CTT
W   R   L   W   K   R   E   K   L   G   K   L   N   A   X
TGG AGG CTC TGG AAA AGG GAA AAG CTG GGC AAA TTG AAT GCC TAA

```

(2) INFORMATIONS FOR SEQ ID NO: 25: ENV PROTEIN (READING FRAME 1)

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: AMINO ACID,

(D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: protéine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```

PKTANLVADITSLAKYQQVLKTLQGTYPXEEGKELFHPCDMVLVKSPLSPNSPSLDTSWEG
PYPVILSTPTAVKVAGVESWIHHTXVKSUILPKEPENPGDNASYSCPLEDLRLLFKQQP
GGKXLKSXIPMALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAP
SYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFT
QTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHRLVSLFNT
TLTGLHEVSAQNPTNCWICLPLNFRPYVSI PVPEQWNNFSTEINTTSVLVGPLVSNLEIT
HTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFVFCGTSAYRCLNGSSSESMCFL
SFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKYL
SQELNGDMERVADSLVTLQDQLNSLA AVLQNRRALDLLTAERGGTCLFLGEECCYYVNQ
SGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFN
LLVNEFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLL
RPNSAGSSXSXGRRPTSPTALRFSCX

```

(2) INFORMATIONS FOR SEQ ID NO: 26: gag PROTEIN

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: AMINO ACID
(D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TSFVEKANGVKCHKYKLSFXHETTHNYVKSVIYALQEAFRVYLPIPASPTPSPTNKDPPS
TQMVQKEIDKRVNSEPKSANIPQLXPLQAVGGREFGPARVHVPFSLPDLKQIKTDLGKFS
DNPBGYIDVLQGLGQFFDLTWRDIMSLLNQTLPNERSATITAAXEFGDLWYLSQVNDRM
TTEEREXFPTGQQAVPSLDPHWDTESEHGDWCCRHLITCVLEGLRKTRKKSMMNYSMMSTI
TQGREENPTAFLERLREALRKRASLSPDSSEGQLILKRKFITQSAADIRKKLQKSAVGPE
QNLETLLNLATSVFYNRDQEEQAEQDKRDXXXKGRFSDHPQASGLWRLWKREKLGKLNAX

(2) INFORMATION FOR SEQ ID NO: 27: env PROTEIN (READING FRAME 1)

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: AMINO ACID,
(D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

LysLeuLeuGlnGluAsnLysGluGlnAlaIleThrLeuGluLysThrGlyAsn...PheTyrProGlnAlaGln
ThrSerGlyIleSerValSerThrSerLeuGlyArgTyrPheHisGlyLeuGlyArgGlyLeuProLeu...Asp
ArgLysGlyProArgGlyAsnLysGlyThrSerSer...AsnAsnSerGlnIleArgThrSerProArgLeuThr
Glu...Gln...ProCysPheProGlyHisSerAsnProGlySerIleProGlyValArgTyrThrIleSerLeu
ThrLeuArgLeuLysAlaThrValLeuArgGluGlyArgGluAsnGlu...AsnThrGlnArgThrSerLysLys
AlaAsnProGlyAsnProProHisMETAlaCysSerValAlaTyrSerLeuLysLysAsnLeuGlnLeuSerPro
LysSerArgThr...ProIleArgAsnAlaValTrpLysAlaLeuHisAsnGln...ProCysAla...ProLys
ThrAlaAsnLeuValAlaAspIleThrSerLeuAlaLysTyrGlnGlnValLeuLysThrLeuGlnGlyThrTyr
Pro...GluGluGlyLysGluLeuPheHisProCysAspMETValLeuValLysSerLeuProSerAsnSerPro
SerLeuAspThrSerTrpGluGlyProTyrProValIleLeuSerThrProThrAlaValLysValAlaGlyVal
GluSerTrpIleHisHisThr...ValLysSerTrpIleLeuProLysGluProGluAsnProGlyAspAsnAla
SerTyrSerCysGluProLeuGluAspLeuArgLeuLeuPheLysGlnGlnProGlyGlyLys...LeuLysSer
...IleProMETAlaLeuProTyrHisIlePheLeuPheThrValLeuLeuProSerPheThrLeuThrAlaPro
ProProCysArgCysMETThrSerSerSerProTyrGlnGluPheLeuTrpArgMETGlnArgProGlyAsnIle
AspAlaProSerTyrArgSerLeuSerLysGlyThrProThrPheThrAlaHisThrHisMETProArgAsnCys
TyrHisSerAlaThrLeuCysMETHisAlaAsnThrHisTyrTrpThrGlyLysMETIleAsnProSerCysPro
GlyGlyLeuGlyValThrValCysTrpThrTyrPheThrGlnThrGlyMETSerAspGlyGlyGlyValGlnAsp

GlnAlaArgGluLysHisValLysGluValIleSerGlnLeuThrArgValHisGlyThrSerSerProTyrLys
GlyLeuAspLeuSerLysLeuHisGluThrLeuArgThrHisThrArgLeuValSerLeuPheAsnThrThrLeu
ThrGlyLeuHisGluValSerAlaGlnAsnProThrAsnCysTrpIleCysLeuProLeuAsnPheArgProTyr
ValSerIleProValProGluGlnTrpAsnAsnPheSerThrGluIleAsnThrThrSerValLeuValGlyPro
LeuValSerAsnLeuGluIleThrHisThrSerAsnLeuThrCysValLysPheSerAsnThrThrTyrThrThr
AsnSerGlnCysIleArgTrpValThrProProThrGlnIleValCysLeuProSerGlyIlePhePheValCys
GlyThrSerAlaTyrArgCysLeuAsnGlySerSerGluSerMETCysPheLeuSerPheLeuValProProMET
ThrIleTyrThrGluGlnAspLeuTyrSerTyrValIleSerLysProArgAsnLysArgValProIleLeuPro
PheValIleGlyAlaGlyValLeuGlyAlaLeuGlyThrGlyIleGlyGlyIleThrThrSerThrGlnPheTyr
TyrLysLeuSerGlnGluLeuAsnGlyAspMETGluArgValAlaAspSerLeuValThrLeuGlnAspGlnLeu
AsnSerLeuAlaAlaValValLeuGlnAsnArgArgAlaLeuAspLeuLeuThrAlaGluArgGlyGlyThrCys
LeuPheLeuGlyGluGluCysCysTyrTyrValAsnGlnSerGlyIleValThrGluLysValLysGluIleArg
AspArgIleGlnArgArgAlaGluGluLeuArgAsnThrGlyProTrpGlyLeuLeuSerGlnTrpMETProTrp
IleLeuProPheLeuGlyProLeuAlaAlaIleIleLeuLeuLeuLeuPheGlyProCysIlePheAsnLeuLeu
ValAsnPheValSerSerArgIleGluAlaValLysLeuGlnMETGluProLysMETGlnSerLysThrLysIle
TyrArgArgProLeuAspArgProAlaSerProArgSerAspValAsnAspIleLysGlyThrProProGluGlu
IleSerAlaAlaGlnProLeuLeuArgProAsnSerAlaGlySerSer...SerGlyArgArgProThrSerPro
ThrAlaLeuArgPheSerCys...AspGlyGly

(2) INFORMATION FOR SEQ ID NO: 28: env protein (open reading frame 2)

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: amino acid,

(D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

SerSerPheArgArgThrLysAsnArgProLeuProTrpArgArgLeuAlaThrAspPheThrHisLysProLys
ProGlnGlyPheGlnTyrLeuLeuValTrpValAspThrPheThrGlyTrpAlaGluAlaPheProCysArgThr
GluLysAlaGlnGluValIleLysAlaLeuValHisGluIleIleProArgPheGlyLeuProArgGlyLeuGln
SerAspAsnSerProAlaPheGlnAlaThrValThrGlnGlyValSerGlnAlaLeuGlyIleArgTyrHisLeu
HisCysAla...ArgProGlnSerSerGlyLysValGluLysMETAsnGluThrLeuLysGlyHisLeuLysLys
GlnThrGlnGluThrHisLeuThrTrpProAlaLeuLeuProIleAlaLeuLysArgIleCysAsnPheProGln
LysAlaGlyLeuSerProTyrGluMETLeuTyrGlyArgProPheIleThrAsnAspLeuValLeuAspProArg

GlnProThr...LeuGlnThrSerProPro...ProAsnIleAsnLysPheLeuLysHisTyrLysGluProIle
ProGluLysArgGluLysAsnTyrSerThrLeuValThrTrpTyr...SerSerProPheProLeuIleProHis
Pro...IleHisProGlyLysAspProThrGlnSerPheTyrLeuProGlnLeuArgLeuLysTrpLeuGluTrp
SerLeuGlyTyrIleThrLeuGluSerAsnProGlyTyrCysGlnArgAsnLeuLysIleGlnGluThrThrLeu
AlaIleProValAsnLeu...ArgIleCysAlaCysSerSerAsnAsnAsnGlnGluGluSerAsn...AsnHis
LysSerProTrpProSerLeuIleIlePhePheSerLeuLeuPhePheTyrProLeuSerLeuSerLeuHisPro
LeuHisAlaAlaVal...ProValAlaProLeuThrLysSerPheTyrGlyGluCysSerValProGluIleLeu
METProHisArgIleGlyValPheLeuArgGluProProProSerLeuProThrProIleCysProAlaThrAla
IleThrLeuProLeuPheAlaCysMETGlnIleLeuIleIleGlyGlnGluLys...LeuIleLeuValValLeu
GluAspLeuGluSerLeuSerValGlyLeuThrSerProLysLeuValCysLeuMETGlyValGluPheLysIle
ArgGlnGluLysAsnMET...LysLys...SerProAsnSerProGlyTyrMETAlaProLeuAlaProThrLys
Asp...IleSerGlnAsnTyrMETLysProSerValProIleLeuAlaTrp...AlaTyrLeuIleProProSer
LeuGlySerMETArgSerArgProLysThrLeuLeuThrValGlyTyrAlaSerPro...ThrSerGlyHisMET
PheGlnSerLeuTyrLeuAsnAsnGlyThrThrSerAlaGlnLys...ThrProLeuProPhe.....AspLeu
LeuPheProIleTrpLys...ProIleProGlnThrSerProVal...AsnLeuAlaIleLeuHisThrGlnPro
ThrProAsnAlaSerGlyGly...LeuLeuProHisLys...SerAlaTyrProGlnGluTyrPheLeuSerVal
ValProGlnProIleValVal...METAlaLeuGlnAsnLeuCysAlaSerSerHisSer...CysProLeu...
ProSerThrLeuAsnLysIleTyrThrValMETSerTyrLeuSerProAlaThrLysGluTyrProPhePheLeu
LeuLeu...GluGlnGluCys...ValHis...ValLeuAlaLeuAlaValSerGlnProLeuLeuSerSerThr
ThrAsnTyrLeuLysAsn...METGlyThrTrpAsnGlySerProThrProTrpSerProCysLysIleAsnLeu
ThrPro...GlnGln...SerPheLysIleGluGluLeu...ThrCys...ProLeuLysGluGlyGluProVal
TyrPhe...GlyLysAsnAlaValIleMETLeuIleAsnProGluSerSerLeuArgLysLeuLysLysPheGlu
IleGluTyrAsnValGluGlnArgSerPheGluThrLeuAspProGlyAlaSerSerAlaAsnGlyCysProGly
PheSerProSer...AspLeu...GlnLeu...TyrCysTyrSerSerLeuAspProValSerLeuThrSerLeu
LeuThrLeuSerLeuProGluSerLysLeu...AsnTyrLysTrpSerProArgCysSerProArgLeuArgSer
ThrAlaAspProTrpThrGlyLeuLeuAlaHisAspLeuMETLeuMETThrSerLysAlaProLeuLeuArgLys
SerGlnLeuHisAsnLeuTyrTyrAlaProIleGlnGlnGluAlaValArgAlaValValGlyGlnProProGln
GlnHisLeuGlyPheProValGluMETGly

(2) INFORMATIONS FOR SEQ ID NO: 29: env protein (open reading frame 3)

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: amino acid,
(D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: protéine
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AlaProSerGlyGluGlnArgThrGlyHisTyrProGlyGluAspTrpGlnLeuIleLeuProThrSerProAsn
LeuArgAspPheSerIleTyr...SerGly...IleLeuSerArgValGlyGlnArgProSerProValGlyGln
LysArgProLysArg.....ArgHis...PheMETLys...PheProAspSerAspPheProGluAlaTyrArg
ValThrIleAlaLeuLeuSerArgProGln...ProArgGluTyrProArgArg...ValTyrAspIleThrTyr
ThrAlaProGluGlyHisSerProGlnGlyArgSerArgLys...METLysHisSerLysAspIle...LysSer
LysProArgLysProThrSerHisGlyLeuLeuCysCysLeu...Pro...LysGluSerAlaThrPheProLys
LysGlnAspLeuAlaHisThrLysCysCysMETGluGlyProSer...ProMETThrLeuCysLeuThrGlnAsp
SerGlnLeuSerCysArgHisHisLeuLeuSerGlnIleSerThrSerSer...AsnIleThrArgAsnLeuSer
LeuArgArgGlyLysArgThrIleProProLeu...HisGlyIleSerGlnValProSerLeu...PheProIle
ProArgTyrIleLeuGlyArgThrLeuProSerHisPheIleTyrProAsnCysGly...SerGlyTrpSerGly
ValLeuAspThrSerHisLeuSerGlnIleLeuAspThrAlaLysGlyThr...LysSerArgArgGlnArg...
LeuPheLeu...ThrSerArgGlyPheAlaProAlaLeuGlnThrThrThrArgArgLysValThrLysIleIle
AsnProHisGlyProProLeuSerTyrPheSerLeuTyrCysSerPheThrLeuPheHisSerHisCysThrPro
SerMETProLeuTyrAspGln...LeuProLeuProArgValSerMETGluAsnAlaAlaSerArgLysTyr...
CysProIleVal...GluSerPhe...GlyAsnProHisLeuHisCysProHisProTyrAlaProGlnLeuLeu
SerLeuCysHisSerLeuHisAlaCysLysTyrSerLeuLeuAspArgLysAsnAsp...Ser...LeuSerTrp
ArgThrTrpSerHisCysLeuLeuAspLeuLeuHisProAsnTrpTyrVal...TrpGlyTrpSerSerArgSer
GlyLysArgLysThrCysLysArgSerAsnLeuProThrHisProGlyThrTrpHisLeu...ProLeuGlnArg
ThrArgSerLeuLysThrThr...AsnProProTyrProTyrSerProGlyLysProIle...TyrHisProHis
TrpAlaPro...GlyLeuGlyProLysProTyr...LeuLeuAspMETProProProGluLeuGlnAlaIleCys
PheAsnProCysThr...ThrMETGluGlnLeuGlnHisArgAsnLysHisHisPheArgPheSerArgThrSer
CysPheGlnSerGlyAsnAsnProTyrLeuLysProHisLeuCysLysIle...GlnTyrTyrIleHisAsnGln
LeuProMETHisGlnValGlyAsnSerSerHisThrAsnSerLeuProThrLeuArgAsnIlePheCysLeuTrp
TyrLeuSerLeuSerLeuPheGluTrpLeuPheArgIleTyrValLeuProLeuIleLeuSerAlaProTyrAsp
HisLeuHis...ThrArgPheIleGlnLeuCysHisIle...AlaProGlnGlnLysSerThrHisSerSerPhe
CysTyrArgSerArgSerAlaArgCysThrArgTyrTrpHisTrpArgTyrHisAsnLeuTyrSerValLeuLeu
GlnThrIleSerArgThrLysTrpGlyHisGlyThrGlyArgArgLeuProGlyHisLeuAlaArgSerThr...

LeuProSerSerSerSerProSerLysSerLysSerPheArgLeuAlaAsnArg...LysArgGlyAsnLeuPhe
IlePheArgGlyArgMETLeuLeuLeuCys...SerIleArgAsnArgHis...GluSer...ArgAsnSerArg
SerAsnThrThr...SerArgGlyAlaSerLysHisTrpThrLeuGlyProProGlnProMETAspAlaLeuAsp
SerProLeuLeuArgThrSerSerSerTyrAsnIleAlaThrProLeuTrpThrLeuTyrLeu...ProProCys
...LeuCysLeuPheGlnAsnArgSerCysLysThrThrAsnGlyAlaGlnAspAlaValGlnAsp...AspLeu
ProGlnThrProGlyProAlaCys...ProThrIle...Cys.....HisGlnArgHisProSer...GlyAsn
LeuSerCysThrThrSerThrThrProGlnPheSerArgLysGlnLeuGluArgSerSerAlaAsnLeuProAsn
SerThr...ValPheLeuLeuArgTrpGly

(2) INFORMATION FORc SEQ ID NO:30 : G1F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30 :

GGACCATAGAGGACACTCCAGGACTA

(2) INFORMATION FOR SEQ ID NO:31 : G1R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31 :

CCTCAGTCCTGCTGCTGGATCATCT

(2) INFORMATION FOR SEQ ID NO:32 : G2F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32 :

CCTCCAAGCAGTGGGAGGAAGAGAATT

(2) INFORMATIONS FOR SEQ ID NO:33 : G2R

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDS NUMBER: single
 - (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33 :

CCTTCCCTGTGTTATTGTGGACATCATT

(2) INFORMATIONS FOR SEQ ID NO:34 : G4F

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDS NUMBER: single
 - (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34 :

GGAAGAAGTCTATGAATTATTCAATGATGT

(2) INFORMATIONS FOR SEQ ID NO:35 : G3F

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDS NUMBER: single
 - (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35 :

GGGACACAGAATCAGAACATGGAGATT

(2) INFORMATIONS FOR SEQ ID NO:36 : G4R

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDS NUMBER: single
 - (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36 :

GCCTTCAGAAGAGTCAGGTGACAGAGA

(2) INFORMATIONS FOR SEQ ID NO:37 : G5R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37 :

GAGCCTCCAAAGTCCACTTGCCTGA

(2) INFORMATIONS FOR SEQ ID NO:38 : E1F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38 :

GATTTTCAGTATCTACTAGTCTGGGTAGAT

(2) INFORMATIONS FOR SEQ ID NO:39 : E1R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39 :

CTAGGAAATCCAGCTAGTCCTGTCTCA

(2) INFORMATIONS FOR SEQ ID NO:40 : E2F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40 :

CCAAGACAGCCAACTTAGTTGCAGACAT

(2) INFORMATIONS FOR SEQ ID NO:41 : E2R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41 :

GGACGCTGCATTCTCCATAGAACTCTT

(2) INFORMATIONS FOR SEQ ID NO:42 : E3F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42 :

GCAATACTACATACACAACCAACTCCCAA

(2) INFORMATIONS FOR SEQ ID NO:43 : E3R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43 :

GGGGGAGGCATATCCAACAGTTAGTA

(2) INFORMATIONS FOR SEQ ID NO:44 : E4F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44 :

CCATCTACTGAACAAGATTTATACTT

(2) INFORMATIONS FOR SEQ ID NO:45 : E4R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45 :

AATGCCAGTACCTAGTGCACCTAGCACT

(2) INFORMATIONS FOR SEQ ID NO:46 : E5F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46 :

CGAATACAACGTAGAGCAGAGGAGCTTCGAA

(2) INFORMATIONS FOR SEQ ID NO:47 : E6F

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleotide
- (C) strands number: single
- (D) CONFIGURATION: line

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47 :

AGCCCAAGATGCAGTCCAAGACTAAGAT

(2) INFORMATIONS FOR SEQ ID NO:48 : E5R

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleotide
- (C) NUMBER OF STRANDS: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48 :

GCGTAGTAGAGGTTGTGCAGCTGAGAT

(2) INFORMATIONS FOR SEQ ID NO:49 : ExF

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49 :

CCCTTACCAAGAGTTTCTATGGAGAAT

(2) INFORMATIONS FOR SEQ ID NO:50 : ExR

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleotide
- (C) STRANDS NUMBER: single
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: cDNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50 :

ACCGCTCTAACTGCTTCCTGCTGAATT

(2) INFORMATIONS FOR SEQ ID NO: 51: gag protein

(i) SEQUENCE CHARACTERISTICS:

- (B) TYPE: amino acid,
- (D) CONFIGURATION: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

TSFVEKANGVKCHKYKLSFHXETTHNYVKSVIYALQEAFRVYLPILPASPTPSPTNKDPPSTQMVQKEIDKRVNSEPKSA
NIPQLXPLQAVGGREFGPARVHVFPFSLPDLKQIKTDLGKFSNPDGYIDVLQGLGQFFDLTWRDIMSLLNQTLPNERSA
TITAAXEFGDLWYLSQVNDRTTEEREXFPTGQQAVPSLDPHWDTESEHGDWCCRHLITCVLEGLRKTRKKSMMNYSMMST
ITQGREENPTAFLERLREALRKRAASLSPDSSEGQLILKRKFITQSAADIRKKLQKSAVGPEQNLETLLNLATSVFYNRDQ
EEQAEQDKRDXXKGHRFSDPQASGLWRLWKREKLGKLNAXXGLLPVRSTRTLXKRLSKXKXAAAPSSMPLISRESLEGPL
PQGTKVLXVRSHXPD/SSSRT

CLAIMS

1. A purified nucleic acid fragment, characterized
5 in that it comprises all or part of a sequence encoding
a human endogenous retroviral sequence, which has at
least *env*-type retroviral motifs, corresponding to the
sequence SEQ ID NO: 1 or to a sequence exhibiting a
10 level of homology with the said sequence SEQ ID NO: 1
greater than or equal to 80% on more than 190
nucleotides or greater than or equal to 70% on more
than 600 nucleotides for the *env*-type domains.
2. The nucleic acid fragment as claimed in
15 claim 1, characterized in that it has retroviral motifs
corresponding to an *env* domain and corresponding to the
sequence SEQ ID NO: 1 and retroviral motifs
corresponding to a *gag* domain and corresponding to the
sequence SEQ ID NO: 2 or to a sequence exhibiting a
20 level of homology greater than or equal to 80% on more
than 190 nucleotides or greater than or equal to 70% on
more than 600 nucleotides for the *env*-type domains and
a level of homology greater than or equal to 90% on
more than 700 nucleotides or greater than or equal to
25 70% on more than 1 200 nucleotides for the *gag*-type
domains, the said motifs having no insertion or
deletion of more than 200 nucleotides.
3. A nucleic acid fragment, characterized in that
it comprises a segment of a sequence as claimed in
claim 1 or claim 2 and in particular the sequence
30 SEQ ID NO: 3-24, the complementary nucleic sequences
and the reverse sequences complementary to the
preceding sequences as well as fragments derived from
the coding regions of the preceding sequences
corresponding to a shifting frame greater than or equal
35 to 14 nucleotides or their complementary sequences.

4. Transcripts, characterized in that they are generated from the sequences as claimed in any one of claims 1 to 3.

5. A diagnostic reagent for the differential
5 detection of complete or partial human endogenous nucleic sequences, having retroviral motifs, selected from the sequences SEQ ID NO: 1 and/or SEQ ID NO: 2, characterized in that it is selected from the group consisting of the sequences SEQ ID NO: 1-50, the
10 complementary nucleic sequences and the reverse sequences complementary to the preceding sequences, of nucleotide fragments capable of defining or of identifying the sequences SEQ ID NO: 1 and/or SEQ ID NO: 2 and any flanking sequence or any sequence
15 overlapping them as well as of fragments derived from the coding regions of the sequences SEQ ID NO: 1-24, corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences, optionally labeled with an appropriate label.

20 6. The reagent as claimed in claim 5, characterized in that it is chosen from the regions situated between nucleotides 3065 and 4390, nucleotides 6965 and 9550 of SEQ ID NO: 3.

7. The reagent as claimed in claim 5,
25 characterized in that it is selected from the sequences SEQ ID NO: 30-50, and in that it is capable of being used as a primer.

8. The reagent as claimed in claim 5,
30 characterized in that it is selected from the following sequences:

- a fragment of 1505 nt amplified by the pair of primers SEQ ID NO: 30 and SEQ ID NO: 31 primers G1F and G1R),
- a fragment of 2529 nt amplified by the pair
35 of primers SEQ ID NO: 38 and SEQ ID NO: 39 (primers E1F and E1R),

and in that it is capable of being used as a probe.

9. A method for the rapid and differential detection of the endogenous retroviral nucleic sequences of the *env* or *env* and *gag* type, their normal or pathological variants, by hybridization and/or gene amplification, carried out using a biological sample, which method is characterized in that it comprises:

(a) a step in which a biological sample to be analyzed is brought into contact with at least one probe as claimed in claim 5, claim 6 or claim 8, and

(b) a step in which the product(s) resulting from the nucleotide sequence-probe interaction is detected by any appropriate means.

10. The method of detection as claimed in claim 9, characterized in that it comprises:

* prior to step (a):

. a step of preparing the relevant biological tissue or fluid,

. a step of extracting the nucleic acid to be detected, and

. at least one gene amplification cycle carried out with the aid of at least one reagent as claimed in any one of claims 5 to 7, and

* subsequent to step (b):

. a step of comparing the nucleic sequences obtained in the said biological sample with the human endogenous retroviral sequences as claimed in any one of claims 1 to 3, by any appropriate means and in particular by sequencing, Southern blotting, restriction cleavage, SSCP or any other method which makes it possible to identify an insertion or a deletion or a single mutation between the various sequences compared.

11. A method of detecting the transcripts as claimed in claim 4, characterized in that it comprises:

- collecting messenger RNAs obtained from control biological samples and from a similar sample collected from patients, and

5 - the qualitative and/or quantitative analysis of the said mRNAs by *in situ* hybridization, by dot-blot, Northern blotting, RNase mapping or RT-PCR, with the aid of a diagnostic reagent as claimed in any one of claims 5 to 8.

10 12. Translational products, characterized in that they are encoded by a nucleotide sequence as claimed in any one of claims 1 to 3.

13. A peptide, characterized in that it is capable of being expressed with the aid of a nucleotide sequence selected from the group consisting of the sequences
15 SEQ ID NO: 1-24 as claimed in any one of claims 1 to 3, according to the combinations offered by usage of the different possible open reading frames.

14. The peptide as claimed in claim 13, characterized in that it includes the derived peptides comprising
20 between 5 and 540 amino acids.

15. The peptide as claimed in claim 13 or claim 14, characterized in that it is selected from:

- . the sequences SEQ ID NO: 25-29 and
- . the sequence SEQ ID NO: 51;

25 .

16. The peptide as claimed in any one of claims 13 to 15, characterized in that it is obtained from nucleic sequences as claimed in any one of claims 1 to 3, in which at least one non-sense codon may be replaced with
30 a codon encoding one of the following amino acids: Phe (F), Leu (L), Ser (S), Tyr (Y), Cys (C), Trp (W), Gln (Q), Arg (R), Lys (K), Glu (E) or Gly (G).

17. An antibody, characterized in that it is directed against one or more of the peptides as claimed in any
35 one of claims 13 to 16.

18. A method for the differential immunological screening of normal or pathological human endogenous

retroviral sequences of the HERV-7q family, characterized in that it comprises bringing a biological sample into contact with an antibody as claimed in claim 17, the reading of the result being
5 visualized by an appropriate means, in particular EIA, ELISA, RIA, fluorescence.

19. A method for the identification and detection of endogenous retroviral motifs which are abnormally expressed in the context of pathological conditions
10 associated with cancer, or of neuropathological conditions, in particular autoimmune neuropathological conditions, at the forefront of which is multiple sclerosis, characterized in that it comprises the comparative analysis of the sequences extracted from a
15 biological sample and the sequences as claimed in any one of claims 12 to 16.

18. An application of the sequences as claimed in any one of claims 1 to 6 or 12 to 16 to the diagnosis of, to the prognosis of, to the evaluation of genetic
20 susceptibility to, any induced, congenital or acquired human diseases, in particular those with cancerous, autoimmune and/or neurological components, such as multiple sclerosis, the associated syndromes and the neurodegenerative diseases in which all or part of the
25 sequences as claimed in to any one of claims 1 to 5 and related endogenous or exogenous forms are involved.

19. Hybrid nucleic sequences, characterized in that they comprise sequences or motifs as claimed in any one of claims 1 to 5, combined with sequences or motifs of
30 endogenous origin or of exogenous origin or induced exogenously.

20. A recombinant cloning or expression vector, characterized in that it comprises a nucleic sequence as claimed in any one of claims 1 to 4.

CCCTGGGGCGGGCTTCCTTTCTGGGATGAGGGCAAAACGGCTGGAGATACAGCAATTATCTTGCAATGAG	74	
AGACAGGACTAGCTGGATTTCTTAGGCGGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACGTCAC	143	
CTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACATAAAATGCTAATTAGGGA	215	
AAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCTGAGAGCACAGCAGGAGGGACAACTATCGGATA	287	repeat
TAAACCCAGGCAATTCGAGCTGGCAACAGCAGCCGCCCTTGGGTCCCTTCCCTTTGATGGGAGCTGTTTTG	359	region
ATGCTATTTCACTCTATTAAATCTTGCAACTGCACCTTCTGGTCCATGTTTCTACGGCTCGAGCTGAGCT	431	
TTTGTCTACCGTCCACCACTGCTGTTGCCACCAGCCGAGACCTGCCGCTGACTCCCATCCCTCTGGATCC	503	R1
GCAGGGTGTCCGCTGTGCTCTGTATCCAGCGAGGGGCCATTGCCGCTCCCAATTGGGGTAAAGGCTTGCCA	575	
TTGTTCTGCACGGCTAAGTGCTGGGTGTTGTTCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTC	647	
TCCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACACACTGGCCCAAGATTCCATCCCTTGGAA	719	
CCGTGAGGCCAAGAATCCAGGTCAGAGAAACAGAGGCTTGCCACCATCTTGGAAAGCGGCTGCTACCATCT	791	
TGGAAAGTGGTTCACCACCATCTTGGGAGCTCTGTGAGCAAGGACCCCGGTAACATTTGGCAACCACGAA	863	
CGGACATCCAAAGTGGTGAATATATTGGACCACTTTCACCTTGCTATTCTGTCTATCCCTTCTTAGAATTG	935	
GAGGAAAATACCGGGCACTTGGCGCAGTTAAAAACGATTAGTGTGGCCACCGGACTTAAGACTCAGGTGT	1007	
GAGGCTATCTGGGAAGGGCTTTCTAACACCCCAACCCCTTCTGGGTGGGGACTTGGTTTGGCTCAAGCC	1079	
AGCTTCCACTTTCAGTTTCTTGGGGAAGCGAGGGCCGACTAGAGGCGAGAAAGCTGTCTCTGAACTCCC	1151	
GGCAGTAGCCGGTTGAGATCATGGTGTAGCCAGAAGTCTCAACAGTCGCCCATGCATGCACCCCTATCTTC	1223	
CTTCTGACCCATACCTCTCTGGTCCCAACCACTTTCTTCAAAGTGTAGCCCAAAATTTCTCTTACCTC	1295	
TGAATATACCTTCTCTGATCCCTGCTCTCTAGTACTATTGGTTCAGACTTCCATTTCTCTAGCAAGTGT	1367	
ATCTCCAAAGGGATCTAAGGAAGCTCTGCGCTGCTCTTAGGCACCTAGGCTATAACCCAGGAGCTCTTAT	1439	
CCCTGGTGTCTCCCTCCCAATTAGGCATACAGCTCTTGACATGGCGAGTTATGTAGGACCCACTCCCCACAC	1511	
CCTTGGCAGGGCCCCAAGTTTGTAAATGGCTGAGGGAAGAGAGACAGAGGAGAGAGAGAGAAATGGAGGA	1583	
GAAAG	1655	
AGAGAGAGTCAAAG	1727	
GGGTAATTTAAACCTGTACTGATAATTGAAGGTCTTCTCTGTGACCCCTATAGCACTCCAACTCCACTTTG	1799	
TGGTCAGTGTAAATTAAGAGCATAGGCGGAAGCACTGAGGCCATTGACAAACCGTAGCTTCCCTATCAAAA	1871	
TCCTTAACCCAGTAACCCGAGATGGACCAAAATGCATTCACTCGGTAGCGCAACTGCTTTGCTAAAAGTAGA	1943	
AAAGTAACCTTTAGAGGAACCACTTATGTGAGCACACCTCACCTGTTGAGAATTATTCTAATAAAAAAGCA	2015	
AAAAGGTAGCTTACTAATCAAAAATCTTAAAGTATGGGGCTATTCTGTTAGAAAAAGGTAATGTAACTCCA	2087	
ACCAGTGATAATTCCTTAACCCAGCAGATTCTTAACGGGATTTAAATCTTAAATACCATACAAAGGTCGG	2159	
ACCAGACCTAGGCGGAATCCCTTCAGGACAGGACGATAGATGGTTCCTCCAGGTGATTGAGGAAAAAAC	2231	
CACAATGGGTATTTCAGTAATTGATACGGGGACTCTTGTGGAAGCAGAGTTAGAAAAAATGCCAATAACTGG	2303	
TCTCTCAAAAGTGTGAGCTGTTTGCACCTCAGCCAGCCCTTAAAGTACTTACAGAATCAAAGACTATCTCA	2375	
ATCCTGATTCAAAAGGTAGCTAGACCCCTCTCTGTAATGCATTGCAATAAGAACTTGTATGGGAATGCAT	2447	
CTTGATGGGGCAGCTGGGTGTTTATAAAATAGGAACCCAGCCAGCTTAGGACTACCCCTGAGCGCAAG	2519	tandem
CGAATGTTGGGCACTGCTGTTAAAGGACCACTAGAATCCAGCAGCCAGCCAGACCCCTTTCTTTGTGGTCAAGAA	2591	repeat
GGCGGGAAGAGGGGTGCAAGGACTGCTACATCGGTAAGCATAAATCCGATAAACAGAGGTCATGGGTGG	2663	regions
TTACGCCACCTGGAAGGAATCACCCTTGAGCACAAGGCAATGTTGGGCAGCGTGGTAAAGGACCACTAG	2735	32
AATCAGCAGCCTGGACCCCTTTCTTTGTGGTCAAGAGAGGCAAGGTAAGGTCAGGACTGCAACATCAG	2807	
TGAGCATAACTAATTGATTAAGCAGAGGTCCATGGGTGGTGTATGCACCTGGAAGAATAAGCATTAGGACC	2879	
ATAGAGGACACTCCAGGACTAAAGCTCATCGGAAATGACTAGGGTTGCTGGCATCCCTATGTTCTTTTC	2951	
AGATGGGAACGTTCCCGCAAGACAAAACGGCCCTAAGACGTATTCTGGAGAATTGGGACCAATTGACC	3023	
CTCAGACACTAAGAAAGAAAGCACTTATATTCTTCTGAGAGTCCCGCTGGCCTGAGGGAGGTATAAT	3095	
TATAACACCATCTTACAGCTAGACCTCTTTTGTAGAAAAGGCAATGGAGTGAAGTCCCATAGTACAACT	3167	
TTCTTTTCAATTAGAGACAATCAAAATATGTAAGAAAGTGTGATTTATGCCCTACAGGAAGCCTTCAGAGT	3239	
GTACTCCCTTATCCAGCAATCCCGACTCTTCCCAACTAATAAGGACCCCTTCAACCCCAATGGTCCA	3311	
AAAGGAGATAGACAAAGGGTAAACAGTGAACCAAGAGTGCCAAATATTCCCCAATTATGACCCCTCCAGC	3383	
AGTGGGAGGAAGAGAAATCCGGCCAGCCAGAGTGCAATGTCCTTTTCTCTCCAGACTTAAAGCAATAAA	3455	
AAAGAGCTTAGGTAATTTCTCAGATAACCCCTGATGGCTATATTGATGTTTACAAGGTTAGGACAAATCTT	3527	
TGATCTGACATGGAGAGATATAATGTCACTGCTAAATCAGACACTAACCCCAATAGAGAGAAGTCCACCAT	3599	33
AAGTCAGCCTGAGAGTTTGGGATCTCTGGTATCTCAGTCAGGTCAATGATAGGATGACAAACAGAGGAAAG	3671	domain
AGAATGATTTCCCAAGGCGCAGGAGGCTTCCAGCTTAGACCCCTCATTTGGGACACAGAAATCAGAACTGG	3743	
AGATTGGTCTCCAGACATTTGCTAACTTGTGTGCTAGAAAGGACTAAGGAAACTAGGAAGAGTCTATGAA	3815	
TTACTCAAATGATGTCACCAATAACACAGGGAAGGGAAGAAATCCTACTGCCCTTTCTGGAGAGACTAAGGA	3887	
GGCATTTAGGAAGCGTGCCTCTCTGTCACTGACTCTTCTGAAAGGCCAACTAATCTTAAAGCGTAAGTTAT	3959	
CACCTCAGTCAGCTGCAGACATTAGAAAAAACTTCAAAAGTCTGCCGTAGGCCCGGAGCAAACTTAGAAAC	4031	
CCATTTGAACCTTGGCAACCTCGGTTTTTTATAATAGAGATCAGGAGGAGCAGGCGGAACAGGCAAAAGGGA	4103	
TTAAAAAAGGCCACCCCTTAGTCATGACCTCAGGCAAGTGGACTTTGGAGGCTCTGGAAGGGGAAA	4175	
GCTGGGCAAAATGAAATGCCATAATAGGGCTTGTCTCCAGTGGGCTTACAAGGAGACTTTAAAAAAGATTGTC	4247	
CAAGTAGAAGTAAGCGGCCCCCTGTCATGCCCCCTATTTCAGGGAACTCACTGGAAGGCCACTGCCCA	4319	
GGGCAAAAGGTTCTCTGATGAGAAGCCACTAACAGATGATCCAGCAGCAGGACTGAGGTTGCCCTGGGG	4391	
AAGCGCATCCCATGCCATACCCCTCACAGAGCCCTGGGTATGCTTGACCATTTGAGGGCCAGGAGGTTGTCT	4463	
CCTGGACACTGCTGCGGCTCTTCTAGTCTTACTCTCTGTCGCGGACAACTGTCTCCAGATCTGTCACTAT	4535	
CTGAGGGGCTGCTAAGACGGGCACTAGATACTTCTCCAGCCACTAAGTTATGACTGGGGAGCTTTAT	4607	
CTTTTTCACATGCTTTTCTAATTATGCTTGAAGCCCCACTACCTTGTTAGGGAGAGACATTCTAGCAAAAG	4679	
CAGGGGCCAATTATACACCTGAACATAGGAGAAGGAACACCCGTTGTTGTCCCTGCTTGAGGAAGGAATTA	4751	
ATCCTGAAGCTCTGGCAACAGGAAGCAATATGGACGAGCAAGAATGCCCGTCTGTTCAAGTTAACTAA	4823	
AGGATTTCCACTCTCTTTCTTCCATACCAAGGCAGTACCCCTCAAGCCCAAGGCCCAACAGGACTCCAAAAGA	4895	
TTGTTAAGGACCTAAAAGCCCAAGGCCCTAGTAAAACATGCAGTAACCCCTGCAGTACTCCAATTTAGGAG	4967	201
TACAGAAACCCAACAGACAGTGGAGGTTAGTGCAAGATCTCAGGATTATCAATGAGGCTGTTGTTCTCTA	5039	domain
AGCCAGCTGTACCTAGCCCTTATAGTCTGCTTTCCCAATACAGAGGAAAGCAGAGTGGTTACAGTCTGG	5111	
ACCTTCAGGATGCTCTTCTGCTATCCCTGTACATCCTGACTCAATCTTGTGCTTTGAGTACTT	5183	

FIGURE 1.1

CAAACCCAACATCTCAACTCACCTGGACTATTTTACCSCAAGGGTTCAGGATAGTCCCATCTATTTGGCC	5253
AGGCATTAGCCCAAGACTTGAAGCAATCTCATACCTGGACACTTGTCTTCGGTAGGTGGATGATTTACTT	5127
TTGGCCGCCCATTCAGAAACCTTGTGCCATCAAGCCACCCCAAGCGCTCTTCAATTTCTCGGTACCTGTGGC	5399
TACATGGTTTCCAAACCAAGGCTCAACTCTGCTCACAGCAGGTTACTTAGGGTAAATTTATCCAAAGGCA	5471
CCAGGGCCCTCAGTGAGGAACACATCCAGCCTATCTGGCTTATGCTCATCCCAAAACCTTAAAGCAACTAA	5543
GGGGATTCTTGGCGTAATAGGTTTCTGCGGAAATGGATTCCAGGTATGGCGAAATAGCCAGGTCATTAA	5615
ATACACTAATTAAGGAACTCAGAAAGCCAAATACCCATTAGTAAGATGGACAACTGAAGTAGAAGTGGCTT	5587
TCCAGGCCCTTAACCCAAGCCCAAGTGTAAAGTTTGGCAACAGGGCAAGACTTTTCTTATATGTCACAGAAA	5759
AAACAGGAATAGCTCTAGGAGTCTTACACAGATCCGAGGGATGAGCTTGCAACCTGTGGCATACCTGACTA	5831
AGGAAATTGATGTAGTGGCAAGGGTTGACCTCATTTGTTTACGGGTAGTGGTGGCAGTAGCAGTCTTAGTAT	5903
CTGAAGCAGTTAAATAATACAGGGAAGAGATCTTACTGTGTGGACATCTCATGATGTGAATGGCATACTCA	5975
CTGCTAAAGGAGACTTGTGGCTGTGACAGCAACTGTTTACTTAAATGTCAGGCTCTATTACTTGAAGGGCCAG	5047
TGCTGCGACTGTGCACTTGTGCAACTCTTAAACCCAGCCACATTTCTTCCAGCAATGAAGAAAAGATAAAR	6119
ATAACTGTCAACAGTAATTTCTCAAACTTATGCCACTCGAGGGGACCTTTTAGAGGTTCTTTGACTGATC	6191
CCGACCTCAACTTGTATCTGATGGAAGTTCCTTTGTAGAAAAAGGACTTCGAAAAGTGGGGTATGCAAGTGG	6263
TCAGTGATAATGGAATACCTTGAAGTAATCCCTCACTCCAGGAAC TAGTGCTCAGGTAGCAGAACTAATAG	6335
CCCTCACTTGGGCACTAGAAATTAGGAGAAAGAAAAAGGGCAATATATATACAGACTCTAAATATGCTTACC	6407
TAGTCTCCATGCCCATGACGAATATGGAAGAAAGGGGAATTCCTAATCTCTGAGAGAACCTATCAAAAC	6479
ATCAGGAAGCCATTAGGAAATTTATTGGCTGTACAGAAACCTAAAGAGGTGGCAGTCTTCACTGCGGGG	6551
GTCTCAGAAAGGAAGGAAGGGGAATAGAAGAGAACTGCCAAGCAGATATTGAAGCCAAAGAGCTGCAA	6623
GGCAGGACCTCTCATTAGAAATGCTTATAAAACAACCCCTAGTATAGGGTAATCCCTTCCGGGAACCAAGC	6695
CCCAGTACTCAGCAGGAGAAACAGAAATGGGGAACCTCAGGAGGACAGTTTCTCCCTCGGGACGGCTAGCC	6767
ACTGAAGAAGGGGAAATACTTTTGGCTGCAACTATCTTGGGAAATTACTTAAACCTTTCATCAAACTTTT	6839
CACCTTAGGCATCGATAGCACCCATCAGATGGCCAAATCATTATTACTGGACAGGGCTTTTCAAACTATC	6911
AAGCAGATAGTCAGGGCCTGTGAAGTGTGCCAGAGAAATATCCCTGCGCTTATCGCAAGCTCTTCAGGA	6983
GAACAAAGAAACAGGCTTGTGCTGGAGAACTGGCAACTGATTTTACCCACAAGGCCAAACCTCAGGGAAT	7055
TTCAATCTCTACTAGTCTGGGTAGATACCTTTCACGGTTTGGGACAGGGCTTCCCTGTAGGACAGAAAAGG	7127
CCCAAGAGGTAATAAAGGCACTAGTTCATGAAATATTCCCAGATTCGGACTTCCCGAGGGCTTACAGAGTG	7199
ACAAATAGCCCTGCTTTCAGGCCACAGTAACCCAGGGAGTATCCAGGGCTTAGGTATACGATATCACTTAC	7271
ACTGGCCCTGAAGGCCAAGCTCTCAGGGAAGGTCGAGAAATGAATGAACCACTCAAGGACA TCTAAAAA	7343
AGCAAAACCCAGGAAACCCACCTCAGATGGCTGCTCTGTGCTTATAGCCTTAAAGAAATCTGCAACTTTC	7415
CCCAAAAGCAGGACTTAGGCCATACGAAATGCTGTATGAAGGCCCTTCAATACCAATGACCTTGTGCTTG	7487
ACCCAAGACAGCCCACTTAGTTTGCAGACATCACTCTTACGCCAAATATCAACAAAGTCTTAAACATTACA	7559
AGGAACCTATCCCTGAGAAGGGGAAAGAACTATTCACCCCTTGTGACATGGTATTAGTCAAGTCCCTTCC	7631
CTCTAAATCCCATTCCTAGATACATCTCTGGGAAGGACCTACCCAGTCAATTTATCTACCCCAACTGCGGT	7703
FAAAGTGGCTGGAGTGGAGTCTTGGATACATCACTTGAAGTCAATCTGGAATAGTCCCAAGGAACCTGA	7775
AAATCCAGGAGCAAACTAGCTATCTCTGTGAACCTTAGAGGATTTGCGCTGCTCTTCAAAACAACACC	7847
AGGAGGAAAGTAACTAAATCATAAATCCCAATGGCCCTCCCTTATCATATTTTCTCTTACTGTTGTTTT	7919
ACCTCTCTTCACTCTCACTGCAACCCCTCCATGCGCTGTATGACCAAGTAGCTCCCTTACCAAGAGTTTCT	7991
ATGGAGAAATGCAGGCTCCCGGAAATATGATGCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCACTT	8063
CATGCCCAACCCATATGCCCCGCACTGCTATCACTCTGCCCCTTTTGCATGCAATGCAAACTACTCATTA	8135
TTGGACAGGAAAAATGATTAATCTAGTTGTCTGAGGAGTCTGAGTCACTGTCTGTTGGACTTACTTCAC	9207
CCAACTGGTATGCTGATGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAAAGAAAGTAACTCTC	9279
CCAACTCAGCCGGGTACATGGCACTCTAGCCCTTACAAAGGACTAGATCTCTCAAAACTACATGAACCCCT	9351
CCGTACCCATACTCGCTGGTAAGCCTATTTAATACCACCTCACTGGCTCCATGAGGTCTCGGCCCAAAA	9423
CGCTACTAATGTTGGATATGCTTCCCTCCCTTGAACCTCAGGCCATATGTTTCAATCCCTGTACCTGAACAAATG	9495
GAACAACCTCAGCAGAGAAATAACACCACTTCCGTTTTAGTAGGACCTCTGTTTCCAACTGGAATAAC	9567
CCATACCTCAAACTCAGCTGCTGTAAATTTAGCAATACATACACAACTCCCAATGCATCAGGTG	9639
GGTAACCTCTCCACACAAATAGTCTGCTACCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCG	9711
TTGTTGAATGGCTCTTCAATCTATGCTTCTCTCACTTTAGTGGCCCTATGACCATCTACACTGA	9783
ACAAGATTTATACAGTTATGTCAATCTAAGCCCGCAAAAGAGTACCCATCTCTCTTTTGTATAGG	9855
CAGGAGTGTAGGTGCTAGGTGCTAGGTGCTGCTTGGCGGTATCAAACTCTACTAGTCTACTTCAAACT	9927
ATCTCAAGAACTAAATGGGGACATGGAACGGTTCGCCGACTCCCTGGTCACTTCAAGATCAACTTAACTC	9999
CCTAGCAGCAGTGTCTTCAAAATCGAAGAGCTTTAGACTTGGTAAACCGTGAAGAGGGGGAACCTGTTT	9071
ATTTTAGGGGAAGAAATGCTGTTTATATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCG	9143
AGATCGAATACAACTGAGAGCAGAGGAGCTTCGAAACACTGGACCTGGGGCTCTCTCAGCCAAATGGATGCC	9215
CTGGATTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATGGTACTCTCTTTGGACCTGTATCTTTAA	9287
CCTCGTTGTTAATTTGCTCTTCCAGAAATGGAAGCTGTAAGAACTACAAATGGAGCCCAAGATGCACTCCAA	9359
GACTAAGATCTACCCGAGACCCCTGGACCCGCTGCTAGGCCACGATCTGATGTTAATGACATCAAAAGGCAC	9431
CCCTCTGAGGAAATCTCAGCTGCACAACTCTACTAGGCCCAATTCAGCAGGAAGCAGTTAGAGCGGTC	9503
TCGGCCAACCTCCCCAACACGACTTAGGTTTCTCTGTTGAGATGGGGACTGAGAGACAGGACTAGCTGGAT	9575
TTCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTG	9647
CAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGGTAAGAA	9719
ATAGCCAATCATCTATTGCTGAGAGCACAGCAGGAGGGAACATGATCGGGATATAACCCAAAGTCTTCGAG	9791
CCGGCAACGGCAACCCCTTTGGGTCCCTCTCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACCTAT	9863
TAAATCTTGAACCTGCACTCTTCTGCTCCATGTTTCTACGGCTTGAGCTGAGCTTTCCGCTCGCCATCCACC	9935
ACTGCTGTTTGGCGCCACCCGAGACCCGCGCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTG	10007
CTCCTGATCCAGCGAGGCAACCCATGCGCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCTCTGCATGGCTA	10079
AGTGCTTGGGTTATCTTAATGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTCTGTGACCCACAG	10151
CTTCTAATAGAGCTATAACACTCACCAGTGGCCCAAGGTTCCATTCTTGAATCCATAAGGCCAAGAACCC	10223
CAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGAGCTCTGTGAGCAAGGACCCCAAGTAACACAAACCA	10295
TGAGGTTGCAATGCAATGGGCCACTAATGGTAGAGCAAGAAACAGAAGGCCCTGGTTCTCGAAGGCATC	10367
AGTAGCTGAAATGCCCTGCCCTGGATGCTCTATTCTAGGTGTTTTCTGCTGAAAGCAGATTAAACCCCTT	10439
GTTCACTTCTCAAGTAGGGCTTCTATTACAGCCCAATCAATCCCAACCCAGATGACAT	10500

FIGURE 1.2

3/15



FIGURE 2

FIGURE 3

IPMALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLLWRMORPGNIDAPSYRSLSKG
 TPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGG
 GVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEV
 SAQNPTNCWICLPNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCV
 KFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSESMCFLSFLVPPMT
IYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDM
ERVADSLVTLQDQLNSLAAVVLQNRRLDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRR
AEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIY
RRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASAGSS

FIGURE 4

- 1) NSLAAVVLQNRRLDLLTAESGGTFLFLEEK
- 2) NSLAAVVLQNRRLDLLTAERGGTCLFLGEEC
- 3) DSLAAVTLQNHQGLDLLTAEGGGLCYFLGEDC
- 4) DSLAAVTLQNHQGLDLLIAEGGGLCTFLGEEC
- 5) DSLAAVTLQNCRGLDLLTAEGGGHYTFLEEEC
- 6) LQNRRLDLLFLKEGGLC
- 7) DSLAKVVLQNRRLDLLTAEQGGICLALQEK

FIGURE 5

TSFVEKANGVKCHKYKLSFHXETHNYVKSIVIYALQEAFRVYLPILPASPTPSPTNKDPPSTQMVQKEIDKRVNS
 EPKSANIPQLXPLQAVGGREFGPARVHVPFSLPDLKQIKTDLGKFSNDPDGYIDVLQGLGQFFDLTWRDIMSLN
 QTLTPNERSATITAAXEFGDLWYLSQVNDRMTTEEREXFPTGQAVPSLDPHWDTESEHGDWCCRHLCTVLEGL
 RKTRKXSMNYSMMSTITQGREENPTAFLERLREALRKASLSPDSSEGQLILKRKFITQSAADIRKKLQKSAVGP
 EQNLETLLNLATSVFYNRDQEEQAEQDKRDXXKGHRFSDHPQASGLWRLWKREKLKGLNAXXGLLPVRSTRTLXK
 RLSKXXAAPSMMPLISRESLEGPLPQGTKVLXVRSHXPD/SSSRT

FIGURE 6

FIGURE 7

7/15

01/ TAAATCCCCATGGCCCTCCCTTATCATATTTTTCT
02/ TAAATCCCC-TGGCCCTCCCTTATCATATTTTTCT
03/ TAAATCCCCATGGCCCTCCCTTATCATATTTTTCT
04/ TAGATCCTCATGGCCCTCC-TTGTATATTTTTTT

01/CTTTACTGTTCTTTTA-CCCTCTTTCACCTCTCACTGCACCCCTCCATGCCGCTGTATGACC
02/CTTTACTGTTCTCTTACCCCCCTTTCACCTCTCACTGCACCCCTCCATGCCACTGCACCCCC
03/CTTTACTGTTCTCTTA-CCCCCTTCTCTCTCACTGCACCCCTCCATGCTGCTGTACAACC
04/CTTTACTGTTCTCTTA-CCCCCTTTCACCTCTCACTGAACCCCTCCATGCCACTGTACTACC

01/AGT-----AGCTCCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGT
02/GTCCATGCCCGTCTCATGCCAGTAGCTCCCCTTAGCAAGAGTTTCTATGGAGAATGCAGCGT
03/AGC-----AGCTCCCCTTACCAAGAGTTTCTATGAAGAATGCGGCTT
04/AGT-----AGCTCCCATTACCAAGAGCTTCTATGGACAATGCGGCTT

01/CCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGC
02/CCCGGAAATATTGATGCCCCATTGTATAGGAGTTTATCTAAGGGAACCCCCACCTTCACTGC
03/CCCAGAAATATTGATGCCCCATCAAATAGGAGTTTACCTAAAGGAAACTCCACCTTCACTGC
04/CCTGGAAATATTGATGACCCATCGTATAGGAGTTTTTCTAAGGGAACCCCCATTTTACCAC

01/CCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATGCATGCAAATACTC
02/CCACACCCATATGCCCCACAAGTCTATACTCTGCCACTCTTTGCATGCATGCAAATACTC
03/CCACACCCATATGCCCCACAAGTCTATACTCTGCCACTCTTTGCATGCATGCAAATACTC
04/CCACACCTATATGACCC-----

01/ATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGTCACTGTCTGT
02/ATTATTGGACAGGAAAAACGATTAATCCCAGTTGTCCTGGAGGACTTGGAG-----
03/ATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAAGACTTGGAGCCACTGTCTGT
04/-----

01/TGGACTTACTTCACCCAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAGA
02/--GACTCACTTCACTCATACCAGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAACAGA
03/CGGACTTACTTCACCCATACTGGTATGTCTGAGGGGGTGGAGTTCAAGATCAGGCAAGAGA
04/-----

01/AAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCCTACA
02/AAAACACATAAAGGAAGTAATCTCCCAACTGACCTGGGTACATAGCACCCCTGGCCCCCTACA
03/AAAACATGTAAAGGAAGTAACCTCCCAACTGACCCGGGTACATAGCACCCCTAGCCCCCTACA
04/-----

01/AAGGACTAGATCTCTCAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTA
02/AAGGACTAGATCTCTCAAACTACATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTA
03/AAGGACTAGATCTCTTAAACTACATGAAACCCTCCATACCCATACTTGCTGGTAAGCCTA
04/-----

01/TTTAATACCACCCCTCACTGGGCTCCATGAGGTCTCGGCCCAAACCCCTACTAACTGTTGGAT
02/TTTAATACCACCCCTGACTGGGCTCCATGAGGTCTCGGCCCAAACCCCTACTAACTGTTGGAT
03/TTTAATACCACCCCTCACTGGGCTCCATGAGGTCTCGGTCCAAACCCCTACTAACTGTTGGT
04/-----

01/ATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAAC
02/GTGCCTCCCCCTGCACCTTAGGCCATACATTTCAATCCCTATACCTGAACAATGGAACAAC
03/GTGCCTCCCCCTGTATTTAGGCCATGCATTTCAATCCCTGTACCTGAACAATGGAACAAC
04/-----TGCACCTCAGGCCATACATTTCAATCCCTGTA-----

FIGURE 8.1


```

01/TCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATA
02/TCAGCACAGAAATAAACACCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATA
03/ACAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATA
-----

01/ACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCA
02/ACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACTCCCA
03/ACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTGTAGACACAACCAACTCCCA
04/-----

01/ATGCATCAGGTGGGTAACTCCTCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTG
02/ATGCATCAGGTGGGTAACTCCTCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTG
03/ATGCATCAGGTGGGTAACTCCTCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTG
04/-----

01/TCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCA
02/TCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCA
03/TCTGTGGTACCTTAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCA
04/-----

01/TTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAA
02/TTCTTAGTGCCCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAA
03/TTCTTAGTGCCCCC-ATGACCATTTACACTGAACAAGATTTATACAATTATGTTGTACCTAA
04/-----

01/GCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTATAGGAGCAGGAGTGCTAGGTGCAC
02/GCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTATTGGAGCAGGAGTGCTAGGCGGAG
03/GCCCCACAACAAAAGAGTACTCATTCTTCCTTTTGTATCGGAGCAGGAGTGCTAGGTGGAC
04/-----

01/TAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAA
02/TAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAACTGTCTCAAGAA
03/TAGGTTCTGGCATTGGCGGTACCACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAA
04/-----

01/CTAAATGGGGACATGGAACGGGTGCGCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTC
02/CTTAAAGGTGACATGGAATGGGTGCGTGATACCCTGGTCACCTTGCAAGATCAACTTAACTC
03/CTCAATGGTGACATGGAATGGGTGCGCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTT
04/-----

01/CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGG
02/CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGG
03/CCTAGCATCAGTAGTCCTTCAAAATTGAAGAGCTTTAGACTTGCTAACCTCTGAAAGAGGGG
04/-----

01/GAACCTGTTTTATTTTTAGGGGAAGAATGCTGTTATTATGTT-----
02/GAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTT-----
03/GAAGCTGTTTTATTTTTAGGGGAAGAATGTTGTTATTATGTTATTTTAGCGGAAGAATGTTGT
04/-----

01/-----AATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACA
02/-----AATCAATCCGGAATCATCACCGAGAAAGTTAAAGAAATTCAGGTCGAATATA
03/TATTATGTTAATCAATCCTGAATTGTCACAGAGAAAGTTGAAGAAATTCGAGATTGAATACA
04/-----

01/ACGTAGAGCAGAGGAGCTTCGAAA-CACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCT
02/ACGTAGAGCAAAGGAGCTGCAAAA-CACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCT
03/ACGTAGAACAGAGGAGCTTCAAAAACACCAGACCCTGGGGCCTCCTCAGCCAATGGATGCCCT
04/-----

```

FIGURE 8.2

9/15

01/ GGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTA
02/ GGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTA
03/ GGATTCTCCCCTTCTTAGGATCTCTAGCAGCTCTAATATTGATACTCCTCTTTGGACCCTGTA
04/ -----

01/ TCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAACTA-----
02/ TCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAATCGAAGCAGTAAACTACAAATCGTTC
03/ TCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTACAAATCGTTC
04/ TCTTTAACCTCCTTGTTAAGCTTGTCTCTTGCAGAATCGAAGCTGTAAACTACAAATGCTTG

01/ --CAAATGGAGCCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCCTG
02/ TTCAAATGGAGCCCCAGATGCAGTCCATGAGTAAAATCTACCACGGACCCCTGGACCGGCCTG
03/ TTCAAATGGAACCCCAGATGAAGTCCATGACTAAGATCTACCGTGGACCCCTGGACCGGCCTA
04/ TTTAAATAGAGCCCCAGATGCAGTCCATGGCTAAGATCTACCACGGACCCCTGGACCGGCCTG

01/ CTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAGCTGCAC
02/ CTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCGAGGAAATCTCAACTGCAC
03/ CTAGCCCATGCTCCAATTGTAATGATATCGAACGCACCCCTCCCGAGGAAATCTCAACTGCAC
04/ CTAGCCCATGCTCTGATGTTGATGACATTGAAGGCACGGCTTCCGAGGAAATCTCAACTGCAC

01/ AACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTTCGTCGGCCAACCTCCCC
02/ AACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCC
03/ AACCCTACTATGCCCCAATTCGCGAGGAAGCAGTTAGACTGGTCGTCAGCCAACCTCCCC

04/ GACCCCTACTACACCCAATTTAGCGGGAAGCAATTAGAGCAGCCTATGGCCACCTCCCC

FIGURE 8.3

CTTCCCCAACTAATAAGGACCCCCCTTTCAACCCAAACAGTCCAAAAGGACATAGACAAAAGGA	3
CTTCCCCAACTAATAAGGACCCCCCTTTCAACCCAAACAGTCCAAAAGGACATAGACAAAAGGA	4
CTTCCCCAACTAATAAGGACCCCCCTTTCAACCCAAATGGTCCAAAAGGAGATAGACAAAAGG	5
CTTCTCCAACTAATAAGGACCCCCCTTTCAACCCAAATGGTCCAAAAGGAGATAGACAAAAGG	6
CTTCCCCAAATAATAAGAACCCCCCTTTCAACCCAAACGGTCCAAAAGGAGATAGACAAAAGG	7
GTAAACAATGAACCAAAGAGTGCCAATATTCCCTGGTTATGCACCCTCCAAGCGGTGGGAG--	3
GTAAACAATGAACCAAAGAGTGCCAATATTCCCTGGTTATGCACCCTCCAAGCGGTGGGAG--	4
GTAAACAGTGAACCAAAGAGTGCCAATATTCCCAATTATGCACCCTCCAAGCAGTGGGAGGA	5
GTAAACAATGAACCAAAGAGTGCCAATATTACACGATTATACTCGCTCCAAGCAGTGGGAG--	6
GTAAACAATAACCAAAGAATGCCAATATTCCCCGATTATGCCCCCTCCAAGCGGTGGGAG--	7
A-AGAATTCGGCCAGCCAGAGTGTCATGTACCTTTTTCTCTCTCAC-ACCTGAAGCAAATTAAA	3
A-AGAATTCGGCCAGCCAGAGTGTCATGTACCTTTTTCTCTCTCAC-ACCTGAAGCAAATTAAA	4
AGAGAATTCGGCCAGCCAGAGTGTCATGTGCCTTTTTCTCTCTCCAG-ACCTAAAGCAAATAAAA	5
-GAGAATTTGGCCAGCCAGCGTGTCATGTACCTTTTTCTCTCTCAG-ATTTAAAGCAAATTAAA	6
-GAGAATTCGGCCAGCCAGAGTGTCACGTACCTTTTTCTCTCTCTAGACTTTAAA----TTAAA	7
ATAGACNTAGGTNAATTNTCAGATAGCCCTGATGGYTATATTGATGTTTTACAAGGATTAGGA	3
ATAGACXTAGGTXAATTTXTCAGATAGCCCTGATGGXTATATTGATGTTTTACAAGGATTAGGA	4
ACAGACTTAGGTAAATTCTCAGATAACCCCTGATGGCTATATTGATGTTTTACAAGGGTTAGGA	5
ATAGACCTAGGTAAATTCTCAGATAACCCCTGATGGCTATATTGATGTTTTACAAGGGTTAGGA	6
ATAGACCTAGGTAAATTCTCAGATAACCCCTAATGGCTATATTGATGTTTTACAAGGTTTAGGA	7
TTCCTGAGTTCTTGCACTAACCTCAAAT	1
CAATCCTTTGATCTGACATGGAGAGATATAATATTACTGCTAAATCAGACGCTAACCTCAAAT	3
CAATCCTTTGATCTGACATGGAGAGATATAATATTACTGCTAAATCAGACGCTAACCTCAAAT	4
CAATCCTTTGATCTGACATGGAGAGATATAATGTCACCTGCTAAATCAGACACTAACCCCAAAT	5
CAATCCTTTGATCTGACATGGAGAGATATAATGTTACTGCTAAATCAGACACTAACCCCAAAT	6
CAATCCTTTGATCTGATATGGAGAGATATAATGTTACTGCTAAATCAGACACTAACCCCAAAT	7
GAGAGAAGTGCCGCCATAACTGCAACCCAAGAGTTTGGCGATCCCTGGTATCTCAGTCAGGTC	1
GAGAGAAGTGCTGCCATAACTGGAGCCCGAGAGTTTGGCAATCTCTGGTATCTCAGTCAGGTC	3
GAGAGAAGTGCTGCCATAACTGGAGCCCGAGAGTTTGGCAATCTCTGGTATCTCAGTCAGGTC	4
GAGAGAAGTGCCACCATAACTGCAGCCTGAGAGTTTGGCGATCTCTGGTATCTCAGTCAGGTC	5
GAAAAAAGTGCTGCCATAACAGCAGCCTGAGAGTTTGGCGAATCTCTGGTATCTCAGTCAGGTC	6
GACAGAAGTGTCGCCGTAACCTGGAGCCCGAGAGTTTGGCAATCTCTGGTATCTCAGTCAGGTC	7
AATGACAGGATGACAACAGAGGAAAGATAATGATTCCCCACAGGCCAGCAGGCAGTTCCCACT	1
AATGATAGGATGACAACGGAGGAAAGAGAACGATTCCCCACAGGGCAGCAGGCAGTTCCCACT	3
AATGATAGGATGACAACGGAGGAAAGAGAACGATTCCCCACAGGGCAGCAGGCAGTTCCCACT	4
AATGATAGGATGACAACAGAGGAAAGAGAATGATTCCCCACAGGCCAGCAGGCAGTTCCCACT	5
AATGATAGGATGACAACAGATGAAAGAGAATGATTCCCCACAGGCCAGCAGGCAGTTCCCACT	6
AATGATAGGATGACAACAGAGGAAAGAGAACGATTCCCCACAGGCCAGCAGGCAGTTCCCACT	7
GTAGACCCTCATTAGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTTGCTAACT	1
AACT	2
GTAGCTCCTCATTGGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTTACTAACT	3
GTAGCTCCTCATTGGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTT	4
CTAGACCCTCATTGGGACACAGAATCAGAACATGGAGATTGGTGCTGCAGACATTTGCTAACT	5
GTAGACCCTCATTAGGACACAGAATCAGAACATGGAGATTGGTGCCACAGACATTTGCTAACT	6
GTAGACCCTCACTGGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTTGCTAACT	7

FIGURE 9.1

TGCGTGCTAGAAAGGACTAAGGAAAAC TAGGAAGA----	1
TGCGTGCTAGAAAGGACTAAGGAAAAC TAGGAAGA---CTATGAATTATTCAATGATGTCCACT	2
TGCGTGCTAGAAAGGACTAAGGAAAAC TAGGAAGA---CTATGAATTATTCAATGATGTCCACT	3
TGTGTGCTAGAAAGGACTAAGGAAAAC TAGGAAGAAGTCTATGAATTACTCAATGATGTCCACA	5
TGCGTGCTAGAAAGGACTAAGGAAAAC TAGGAAGAAGCCCATGAATTATTCAATGATGTCCCCT	6
TGCGTGCTAGAAAGGACTAAGGAAAAC TAGAAAGAAGCCTGTGAGTTATTCAATGATGTCCACT	7
ATAACACAGGGGAAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	1
ATAACACAGGGGAAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	2
ATAACACAGGGGAAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	3
ATAACACAGGG-AAGGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	5
ATAACACAGGG-AAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAAGGATTGAG	6
ATAACACAGGG-AAAGGAAGAAAATCCTACCGCCTTTCTGGAGTGACTAACGGAGGCATTGAG	7
GAAGCATACC---AGGCAAGTGGACATTGGAGGCTCTGGAAAAGGGAAAAGTTGGGAAAAGTA	1
GAAGCATACC---AGGCAAGTGGACATTGGAGGCTCTGGAAAAGGGAAAAGTTGGGCAAATTG	2
GAAGCATACC---AGGCAAGTGGACATTGGAGGCTCTGGAAAAGGGAAAAGTTGGGCAAATTG	3
GAAGCGTGCC232AGGCAAGTGGACTTTGGAGGCTCTGGAAAAGGGAAAAGCTGGGCAAATTG	5
GAAGCATACC238AGGCAAATGGACTTTGGAGGCTCCAGAAAAGGGAAAAGCTGAGCAAATTG	6
GAAGCATACC233AGGCAAGCGGACTTTGGAGGCTCTGGAAAAGGGAAAAGCTAGGCAAATCA	7
TATGTCTAATAGGGCTTGCTTCCAGTGTGGTCTACAAGGACACTTTAAAAAAGATTGTCC-AA	1
AATGCCTAATAGGGCTTGCTTCCAGTGCAGTCTACAAGGACGCTTTAGAAAAGATTGTCC-AA	2
AATGCCTAA	3
AATGCCTAATAGGGCTTGCTTCCAGTGCAGTCTACAAGGACACTTTAAAAAAGATTGTCC-AA	5
AATGCCTAACAGGGCTTGCTTCTAGTGTGGTCTACAAGGACACTTTAAAAAAGATTGTCC-AA	6
AATGCCTAATAGGGTTTGCTTCCAGTGCAGTCTACAAGGACACTTTAAAAAAGATTGTCCAAA	7
-TAGAAATAAGCCACCACCTCGTCCATGCCCCCTTATGTCAAGGGAATCACTGGAAGGCCCACT	1
GTAGAAATAAGCCGCCCC-TCGTCCATGCCCCCTTATGTCAAGGGAATCACTGGAAGGCCCACT	2
GTAGAAGTAAGCCGCCCCCTCGTCCATGCCCCCTTATTTCAAGGGAATCACTGGAAGGCCCACT	5
GTAGAAACAAGCTGCCCCCTTGCTCCATGCCCCCTTATGTCAAGGGAATCACTGGAAGGCCCACT	6
-TAGAAATAAGCCGCCCCCTCGTCCATGCACCTCGTGTCAAGGGAATCACTGTAAGGCCCACT	7
GCCCCAGGGGATGAAGGTCCTCTGAGTCAGAAGCCACTAACCAGATGA	1
GCCCCAGGGGACGAAGGTCCTCTGAGTCAGAAGCCACTAACCTGATGA	2
GCCCCAGGGGACAAAGGTCCTCTGAGTCAGAAGCCACTAACCAGATGA	5
GCCCCAGGAGATGAAGGTCCTCTGAGTCAGAAGCCACTAACCAGATAA	6
GCCCCAGGGGACGTAGGTCCTCTGAGTCAGAAGCCACTAACCAGATGA	7

FIGURE 9.2

10

FIGURE 10

FIGURE 10

FIGURE 11

FIGURE 12

1

FIGURE 13